Digital Signal Processing 40 (2015) 1–30

Contents lists available at ScienceDirect

Digital Signal Processing

www.elsevier.com/locate/dsp

Time–frequency features for pattern recognition using high-resolution TFDs: A tutorial review

Boualem Boashash [a,b,*], Nabeel Ali Khan [a], Taoufik Ben-Jabeur [a]

[a] Qatar University, Department of Electrical Engineering, Doha, Qatar
[b] University of Queensland, Center for Clinical Research, Brisbane, Australia

ARTICLE INFO

Article history:
Available online 8 January 2015

Keywords:
Quadratic time–frequency distributions
Time–frequency image processing
Time–frequency feature extraction
Compact kernel distribution
Newborn EEG seizure abnormality
Non-stationary signals

ABSTRACT

This paper presents a tutorial review of recent advances in the field of time–frequency (t, f) signal processing with focus on exploiting (t, f) image feature information using pattern recognition techniques for detection and classification applications. This is achieved by (1) revisiting and streamlining the design of high-resolution quadratic time frequency distributions (TFDs) so as to produce adequate (t, f) images, (2) using image enhancement techniques to improve the resolution of TFDs, and (3) defining new (t, f) features such as (t, f) flatness and (t, f) entropy by extending time-domain or frequency-domain features. Comparative results indicate that the new (t, f) features give better performance as compared to time-only or frequency-only features for the detection of abnormalities in newborn EEG signals. Defining high-resolution TFDs for the extraction of new (t, f) features further improves performance. The findings are corroborated by new experimental results, theoretical derivations and conceptual insights.

1. Introduction

1.1. Non-stationary signals

There has been a significant effort in the past decades to develop Digital Signal Processing (DSP) methods that are more appropriate for the representation, analysis and processing of non-stationary signals [1]. The studies initially focused on the design of Spectrograms, filter-banks and quadratic methods related to the Wigner-Ville Distribution (WVD) for optimal representation of signals such as mono-component Linearly Frequency Modulated (LFM) signals used in radar [2], sonar [3], geophysics [4], and other applications [5–8]. Researchers then shifted their focus to the analysis of multi-component signals, resulting in a number of new approaches to design advanced Time-Frequency Distributions (TFDs) and related techniques [9–14,8]. Such multi-component signals provide a more precise modeling of most natural signals, including biomedical signals such as fetal movement signal, phonocardiogram (PCG), electrocardiogram (ECG), speech, electroencephalogram (EEG), heart rate variability (HRV) and many others. A typical model for such signals was originally proposed in ([15], p. 419). Following this approach, a multi-component signal of time duration T and bandwidth B can be expressed as:

$$s(t) = \sum_{k=1}^{N_c} s_k(t) = \sum_{k=1}^{N_c} a_k(t) \cos\big(\phi_k(t)\big), \tag{1}$$

where N_c is the total number of components, $s_k(t)$ is the kth signal component, $a_k(t)$ and $\phi_k(t)$ are the instantaneous amplitude (IA) and instantaneous phase (IP) of the kth signal component. The IA $a_k(t)$ is a low-frequency amplitude whose spectrum is assumed not to overlap with the spectrum of the higher-frequency signal $\cos(\phi_k(t))$ [16,17]. The derivative of the IP $f_k(t) = \frac{1}{2\pi}\frac{d\phi_k(t)}{dt}$ defines the instantaneous frequency (IF) of the kth signal component. The above defines the AM-FM model of the signal where the IA is the AM law and the IF is the FM law of the signal. The time support of $a_k(t)$ is T and the frequency support of $f_k(t)$ is B.

Using the above model, a multi-component signal is completely characterized by the number of components N_c, the IA and IF of each component; residual noise could be added in certain situations when there are interferences. The time–frequency (t, f) approach is a preferred method to represent such multi-component signals as it allows solving the problem of estimating the characteristics of the components constituting the signal in Eq. (1). For example, it can provide a measure of the energy leakage that takes place around the components due to slow variation of their IAs, and a measure of the relative energy distribution throughout the (t, f) plane. Such information can then be used to refine a number

* Corresponding author at: Qatar University, Department of Electrical Engineering, Doha, Qatar.
E-mail addresses: boualem@qu.edu.qa (B. Boashash), nabeelalikhan@qu.edu.qa (N.A. Khan), taoufik@qu.edu.qa (T. Ben-Jabeur).

http://dx.doi.org/10.1016/j.dsp.2014.12.015
1051-2004/

of advanced techniques and methodologies in applications such as detection and classification using selected (t, f) features for pattern recognition. An important generic application of such methods is diagnosis, which is related to condition monitoring and fault detection. This applies to both machines and humans. For the sake of illustration, we will use physiological signals in this paper, although the method can be applied to all fields where decisions need to be made on the basis of information collected from non-stationary signals.

1.2. Pattern recognition for diagnosis, condition monitoring and fault detection

The basic idea is to find and recognize patterns in the signal that would indicate the presence of a fault, an abnormality or otherwise would indicate a normal condition for the process generating the signals. If there is an abnormality, then one needs to identify and classify the abnormality. Although the approach is general, this paper illustrates it by focusing on some particular physiological signals such as electroencephalography (EEG) for monitoring new-born brain activity and predicting newborn health outcomes through diagnosis and prognosis.

1.2.1. Acquisition and representation of real signals

Although the methodologies presented are general, this paper considers only physiological signals for illustration, including EEG, electrocardiography (ECG), heart rate variability (HRV) and heart sounds which are routinely used by medical specialists for diagnosis in a range of situations. After acquisition, these signals are usually processed in several stages, including a pre-processing stage like amplification and filtering to suppress artifacts and noise. In some situations, multi-channel physiological signals are acquired using multiple spatially distributed sensors that provide both temporal and spatial information. The diversity given by space time information can allow a more refined analysis and better interpretation of the data in applications such as source localization, signal-to-noise ratio (SNR) enhancement, classification, and removal of artifacts [18,19]. As indicated earlier, (t, f) analysis is widely used for the representation and analysis of physiological signals because of the non-stationary and multi-component characteristics of these signals; the (t, f) image allows a refined treatment of further processing stages such as detection and classification, leading to diagnosis, condition monitoring and fault detection.

1.2.2. Automated TF detection of abnormalities in non-stationary signals

Early detection of abnormalities in physiological signals can help save or improve patients' lives. Manual detection of abnormality requires a constant monitoring of the relevant physiological signal by a qualified medical expert. The development of an automatic abnormality detection technique, which can be implemented on a digital computer, would result in a major advance in medical practice. Such detection of abnormality in physiological signals can be considered as a pattern classification problem involving extraction of features from the signals and training of a classifier. Features can be extracted from the time-domain representation [20–22], the frequency-domain representation [23], a combination of both time-domain and frequency-domain representations [24], or joint time frequency representations using TFDs [7,25,5,26–30,13,31]. The performance of abnormality detection techniques depends of course on the choice of signal representation and selection of features. The abnormality detection techniques that are based on TFDs are shown to outperform time-domain or frequency-domain only approaches as TFDs are more adapted to analyze the non-stationary characteristics of physiological signals [26] and convey critical information about the signal.

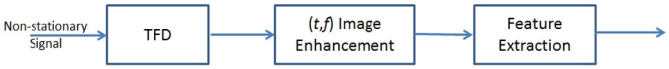


Fig. 1. The proposed (t, f) pattern recognition approach.

TFDs are normally compared in terms of their resolution and cross-term suppression properties. Previous studies have shown that high-resolution TFDs, like the Modified B-Distribution (MBD), give good classification results [25,26]. A key question is therefore whether the performance of TFD-based signal classification techniques can be related to the resolution of the TFD. This paper aims to answer this question by investigating the influence of resolution parameters on classification performance.

TFDs are rich in information partly because they increase the dimensionality of the classification problem at the expense of increased computational costs. In order to avoid this problem, it is important to extract only the most relevant features from the TFD. Such features can be extracted in a number of ways. These include dividing a TFD into a number of tiles and computing the energy in each tile [6], reducing the dimension of the TFD by linear transformation methods [5,30,31], selecting relevant points from a TFD based on relevance or mutual information measure [29], translating time-domain or frequency-domain features to (t, f) domain and extracting relevant features [26], and using image processing techniques to extract features like statistical measures, texture related features, geometric features, ridges [25,28], to name just a few.

1.3. Scope and key topic covered

This study reviews previous work on the design of high-resolution TFDs [32,10], (t, f) image processing and feature extraction methodology [25,26] with three objectives: (1) streamlining the design of high-resolution TFDs, (2) improve diagnostic applications and (3) providing the reader with an illustration on physiological signal classification techniques. The study refines, updates and extends previous studies with improved precision and illustrates the improvements using a real application: enhancing newborn health outcomes. The style of this paper is a tutorial review with scope and coverage defined by the inclusion of the following:

1. A streamlined review of quadratic (t, f) distributions (QTFDs) formulations;
2. A methodology for the design of new high-resolution QTFDs and the study of how they improve the performance of features extraction;
3. A review of multi-component IF estimation techniques as a performance measure to compare TFDs;
4. A review of (t, f) image processing techniques for cross-term suppression, resolution enhancement and de-noising as a pre-processing stage before classification; including the design of a (t, f) image enhancement technique based on directional filtering to improve the resolution of QTFDs;
5. A formulation of (t, f) translated features from time-domain only and frequency-domain only features;
6. Illustrative experiments to apply the above points to EEG seizure detection using a large medical database.

This study develops a (t, f) pattern recognition methodology to link design of high-resolution TFDs, (t, f) image enhancement techniques, and feature extraction stages as illustrated in Fig. 1. Note that the authors have restricted to the review of only QTFDs because of their high resolution, simple interpretation as a distribution of signal energy in a (t, f) plane and widespread use. The rest of the paper is organized as follows. The design of high-resolution QTFDs is discussed in Section 2. Multi-component IF

estimation techniques are described in Section 3 and compared in terms of performance. Then, Section 4 presents a review of image processing techniques for enhancement of TFDs. The methodology for extracting (t, f) features is described in Section 5. Experimental results are described and discussed in Section 6. Finally, Section 7 concludes the paper. Appendix A describes the resolution measure used in this study to compare the performance of TFDs. The issues related to computational cost and real time implementation are discussed in Appendix B. In addition, the programs that are used in this study are described in Appendix C. Appendix D gives a list of (t, f) features obtained by extending time-only features to the (t, f) domain. Appendix E refers the reader to the real data used in this study which appear as Supplementary material.

2. Quadratic time–frequency signal processing

2.1. Time frequency formulations and representations

Let us consider a non-stationary real signal $s(t)$ as defined in Eq. (1). The WVD is the core member of the class of QTFDs. The WVD is defined by taking the Fourier transform (FT) of an instantaneous auto-correlation function $K_z(t, \tau)$ expressed as [1]:

$$W_z(t, f) = \int_{\mathbb{R}} K_z(t, \tau) e^{-2j\pi f \tau} \mathrm{d}\tau, \tag{2}$$

where $K_z(t, \tau)$ is defined as

$$K_z(t, \tau) = z\left(t + \frac{\tau}{2}\right) z^*\left(t - \frac{\tau}{2}\right), \tag{3}$$

and where $z(t)$ is the analytic associate of a real signal $s(t)$ obtained by the use of the Hilbert transform; expressed as

$$z(t) = s(t) + j\mathcal{H}\{s(t)\}. \tag{4}$$

In the above expression, the imaginary part represents the Hilbert Transform defined by $\mathcal{H}\{s(t)\} = \frac{1}{\pi}\text{p.v.}\{\int_{\mathbb{R}} \frac{s(\tau)}{t-\tau}\mathrm{d}\tau\}$, where p.v. represents a principle value expressed as

$$\text{p.v.}\left\{\int_{\mathbb{R}} \frac{s(\tau)}{t-\tau}\mathrm{d}\tau\right\} = \lim_{\delta\to 0}\left[\int_{-\infty}^{t-\delta} \frac{s(\tau)}{t-\tau}\mathrm{d}\tau + \int_{t+\delta}^{+\infty} \frac{s(\tau)}{t-\tau}\mathrm{d}\tau\right]. \tag{5}$$

For a real AM-FM signal $s(t) = a(t)\cos(\phi(t))$, the analytic signal $z(t)$ can be written as: $z(t) = a(t)e^{j\phi(t)}$, under some conditions such as when $a(t)$ is a low-frequency amplitude whose spectrum does not overlap with the spectrum of the high-frequency signal $e^{j\phi(t)}$ [16,17]. For such signals with large Bandwidth–Time-duration (BT) product, then the IF provides an estimate of the FM law.

2.1.1. Direct quadratic time–frequency formulation

The WVD defined in Eq. (2) gives ideal concentration for mono-component LFM signals, but it produces undesired cross-terms for non-linear frequency modulated (FM) or multi-component signals. Cross-terms can be reduced by convolving the WVD with a 2D (t, f) kernel, resulting in the following expression [1],

$$\rho(t, f) = \gamma(t, f) \underset{(t,f)}{**} W_z(t, f) \tag{6}$$

where $\gamma(t, f)$ is the 2D (t, f) smoothing kernel. Eq. (6) represents the (t, f) formulation of QTFDs. The 2D smoothing of the WVD with $\gamma(t, f)$ reduces the cross-terms, but blurs the auto-terms. For this reason, the (t, f) kernels are designed to achieve the best tradeoff between the following two conflicting objectives:

a) Minimize cross-terms;
b) Retain the resolution of auto-terms.

The selected compromise for this trade-off selection depends on the signal specifications. High-resolution TFDs are needed for accurate estimation of signal parameters such as the IF, total number of components which are important features for a number of applications including signal modeling, analysis and classification. The estimation of QTFDs based on the general (t, f) formulation requires one FT and a two dimensional (2-D) convolution along time and frequency axis. Along with this general (t, f) formulation, a QTFD can be expressed starting from a time–lag formulation, Doppler–lag formulation or a Doppler–frequency formulation, as outlined in the next sections.

2.1.2. Time–lag formulation of QTFDs

Eq. (6) can be estimated via time–lag formulation by replacing the convolution operation along the frequency axis with the multiplication operation along the lag axis ([33], p. 67), resulting in the following expression:

$$\rho(t, f) = \int_{\mathbb{R}} \underbrace{G(t, \tau) \underset{t}{*} K_z(t, \tau)}_{R_z(t,\tau)} e^{-2j\pi f \tau} \mathrm{d}\tau, \tag{7}$$

where $G(t, \tau)$ is called the time–lag kernel of the TFD and is related to $\gamma(t, f)$ by inverse FT. The time–lag formulation is widely used for implementing TFDs because of its conceptual simplicity and computational efficiency as it requires only one convolution along the time axis and one FT from lag domain to frequency domain [13].

2.1.3. Doppler–frequency formulation of QTFDs

A Doppler–frequency based formulation of TFDs can be obtained from the (t, f) formulation by replacing the time convolution in Eq. (6) by a multiplication in Doppler after taking a FT [33]:

$$\rho(t, f) = \int_{\mathbb{R}} \mathcal{G}(\nu, f) \underset{f}{*} k_z(\nu, f) e^{2j\pi\nu t} \mathrm{d}\nu, \tag{8}$$

where $\mathcal{G}(\nu, f)$ is the kernel in the Doppler–frequency domain. The quantity $k_z(\nu, f)$ is often referred to as the spectral autocorrelation function (SAF) defined as ([33], p. 70):

$$k_z(\nu, f) = Z\left(f + \frac{\nu}{2}\right) Z^*\left(f - \frac{\nu}{2}\right), \tag{9}$$

where $Z(f)$ is the FT of $z(t)$. The Doppler–frequency TFD formulation requires one convolution along the frequency axis and one inverse FT (from Doppler to time) to transform the SAF to the (t, f) domain representation. The computational cost of transforming the SAF to the (t, f) domain is equal to that of transforming the instantaneous auto-correlation function to the (t, f) domain, apart from the required estimation of $Z(f)$.

2.1.4. Doppler–lag formulation of QTFDs

A Doppler–lag formulation can be obtained by replacing the convolution operation along the time axis in Eq. (7) with the equivalent multiplication along the Doppler-axis ([33], p. 69), resulting in the following expression:

$$\rho(t, f) = \int_{\mathbb{R}} \int_{\mathbb{R}} \underbrace{g(\nu, \tau) A_z(\nu, \tau)}_{\mathcal{A}_z(\nu,\tau)} e^{-2j\pi f\tau + 2j\pi t\nu} \mathrm{d}\tau \mathrm{d}\nu, \tag{10}$$

where $A_z(\nu, \tau)$ represents the ambiguity function, defined as:

$$A_z(\nu, \tau) = \int_{\mathbb{R}} K_z(t, \tau) e^{-2j\pi t\nu} \mathrm{d}t, \tag{11}$$

and $g(\nu, \tau) = \int_{\mathbb{R}} G(t, \tau) e^{-2j\pi t\nu} \mathrm{d}t$ is the Doppler–lag kernel. This formulation is often used for designing high-resolution TFDs

as it allows entering filter specifications in the formulation of $g(\nu,\tau)$ [33]. The Doppler–lag formulation requires three FTs as computing the ambiguity function from the time–lag kernel needs one FT and transforming the ambiguity function to the TFD requires two FTs.

These four formulations are illustrated in Fig. 3.

The selection of one of the above four formulations depends on the specifications of the application and signal characteristics such as duration T, bandwidth B, and value of BT product. For example, for narrow-band signals (i.e. with small B), Eq. (8) would be preferable to use ([33], p. 70).

2.1.5. Which TFD to use? Spectrogram as a non-separable QTFD, WVD, S-method or another TFD?

A question asked by most beginners is which TFD to use to get started, given that the WVD has a good resolution with a problem of cross-terms, the Spectrogram has a problem of resolution, but no apparent cross-terms and other QTFDs appear more complicated. To answer this question, let us first review the Spectrogram in detail.

The Spectrogram is a simple yet effective TFD in some situations. It is defined as:

$$S_s^w(t,f)=\left|\int_{\mathbb{R}} s(\tau)w(t-\tau)e^{-2j\pi f\tau}d\tau\right|^2, \tag{12}$$

where $w(t)$ is the analysis window. The Spectrogram is a QTFD as its time–lag kernel can be expressed as the instantaneous autocorrelation function of the window ([33], pp. 47–48).

$$G(t,\tau)=w^*\left(t+\frac{\tau}{2}\right)w\left(t-\frac{\tau}{2}\right). \tag{13}$$

This implies that the (t,f) kernel of the Spectrogram $\gamma(t,f)$ is the WVD of the window $w(t)$, and equivalently the ambiguity domain kernel filter $g(\nu,\tau)$ is the ambiguity function of the window $w(t)$ ([33], p. 76),

$$\gamma(t,f)=\int_{\mathbb{R}} w^*\left(t+\frac{\tau}{2}\right)w\left(t-\frac{\tau}{2}\right)e^{-2j\pi f\tau}d\tau. \tag{14}$$

To discuss the limitations of the smoothing kernel of the Spectrogram, let's consider a Gaussian window $w(t)=(a/\pi)^{1/4}e^{-\frac{a}{2}t^2}$. Using Eq. (14) results in the (t,f) kernel for this window expressed as [34]:

$$\gamma(t,f)=2e^{-at^2}e^{-\frac{4\pi^2f^2}{a}}. \tag{15}$$

The above expression shows that the smoothing kernel of the Spectrogram is non-separable given that both t and f variables are parametrized by the same scale value. The smoothing along the time axis cannot be controlled independently from the one in the frequency axis and vice-versa; this makes the Spectrogram extremely sensitive to the length of the window ([33], p. 221). Attempts to mitigate this limitation of the Spectrogram include using separable kernel TFDs that have the flexibility to independently adjust the type of smoothing along the time or frequency axis as discussed in Section 2.2.

In short, the Spectrogram is an easy to use QTFD that does not have a problem with cross-terms, but suffers from poor resolution as well as sensitivity to the choice of window length. On the other hand, the WVD has higher resolution for mono-component LFM signals, but has an issue of cross-terms that may be problematic for analysis and modeling.

The S-method, which can be considered an optimized version of the Spectrogram, combines the advantages of the Spectrogram and WVD. It is defined as [35]:

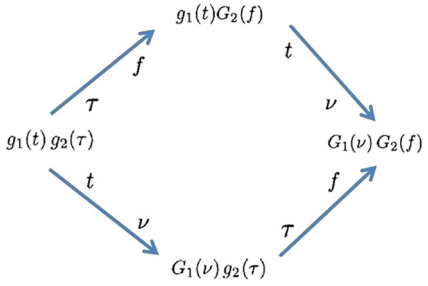


Fig. 2. The relationships between equivalent formulations of separable (t,f) kernels in several dual domains.

$$\mathrm{SM}_s(t,f)=2\int_{\mathbb{R}} G(\theta)F_s(t,f+\theta)F_s^*(t,f-\theta)d\theta, \tag{16}$$

where $F_s(t,f)$ is the short-time Fourier transform of $s(t)$, defined as:

$$F_s(t,f)=\int_{\mathbb{R}} w(\tau)s(t+\tau)e^{-j2\pi f\tau}d\tau. \tag{17}$$

In Eq. (16), $G(\theta)$ is a narrow window whose length controls the cross-term suppression and auto-term resolution properties of the TFD. Previous studies have shown that an appropriate selection of the length of $G(\theta)$ can combine advantages of both Spectrogram and WVD. The resultant TFD can have auto-term resolution close to the WVD with a significant suppression of cross-terms [35]. The ambiguity domain kernel of the S-method is given below [13].

$$\begin{aligned} g(\nu,\tau)&=G(\nu)\underset{\nu}{*}\int_{\mathbb{R}} w\left(u+\frac{\tau}{2}\right)w^*\left(u-\frac{\tau}{2}\right)e^{-j2\pi u\nu}du\\ &=G(\nu)\underset{\nu}{*}A_w(\nu,\tau). \end{aligned} \tag{18}$$

Eq. (16) shows that the S-method is computed using the short-time Fourier transform, which makes it more computationally efficient than other QTFDs such as the WVD. Let us now review the design of separable kernel TFDs.

2.2. High-resolution TFDs design based on separable kernels

2.2.1. Definition of separable kernel TFDs

An attempt at simplification is to first consider (t,f) kernels that can be represented simply as the product of an independent time-only kernel with an independent frequency-only kernel; these separable (t,f) kernels ([33], pp. 213–222) are expressed as:

$$\gamma(t,f)=g_1(t)G_2(f). \tag{19}$$

The meaning and significance of this simplification is that we can define and design some QTFDs simply by smoothing the WVD in t and then in f. The shape and size of $g_1(t)$ or $G_2(f)$ determines the smoothing along the time or frequency axis respectively. In many applications, separable (t,f) kernels have shown to give good enough results in terms of cross-term suppression and auto-term resolution, making them popular as they are easy to use. Other equivalent formulations of (t,f) separable kernels are given below in other related domains.

- Time–lag domain: $G(t,\tau)=g_1(t)g_2(\tau)$
- Doppler–frequency domain: $\mathcal{G}(\nu,f)=G_1(\nu)G_2(f)$
- Doppler–lag domain: $g(\nu,\tau)=G_1(\nu)g_2(\tau)$

Fig. 2 illustrates the relationships between the above formulations of the (t,f) kernels.

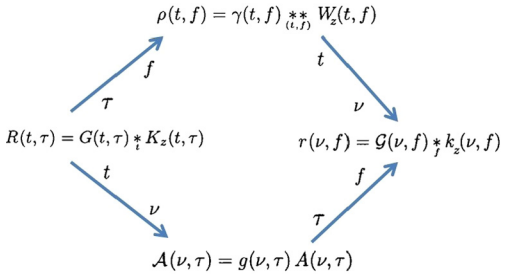


Fig. 3. The relationships between equivalent formulations of TFD in several dual domains.

As discussed above, high-resolution separable kernel TFDs can be designed in the four possible domains, that is: (t, f) domain, Doppler–frequency domain, time–lag domain or Doppler–lag domain; they are related to each other by FT as illustrated in Fig. 3. The ambiguity domain (lag–Doppler domain) is usually preferred for designing TFDs as filtering in the ambiguity domain is performed by a multiplication of lag and Doppler windows instead of one or two convolutions in other domains.

2.2.2. Cross-terms in the WVD and ambiguity domain

The WVD defined in Eq. (2) is often considered as the core distribution of QTFDs; it suffers from the presence of undesired artifacts or cross-terms that prevent its straightforward interpretation as an energy distribution versus t and f. These cross-terms can be classified into two types: outer-terms and inner-terms. Outer-terms can be easily explained by considering a two component signal: $z(t) = z_1(t) + z_2(t)$ for which the WVD is given by

$$W_z(t,f) = W_{z_1}(t,f) + W_{z_2}(t,f) + \underbrace{2\mathrm{Re}\left\{\int_{\mathbb{R}} z_1\left(t+\frac{\tau}{2}\right) z_2^*\left(t-\frac{\tau}{2}\right) e^{-2j\pi f\tau}\,\mathrm{d}\tau\right\}}_{W_{z_1 z_2}(t,f)}. \quad (20)$$

In the above expression, $W_{z_1}(t,f)$ and $W_{z_2}(t,f)$ represent auto-terms that describe an energy concentration while $W_{z_1 z_2}(t,f)$ represents outer-term artifacts that are generated by the interaction of signal components $z_1(t)$ and $z_2(t)$. The second type of artifacts is called inner-terms; they are generated by the non-linear FM characteristic of mono-component signals.

A non-linear mono-component signal can be approximated by the summation of a number of piece-wise LFM signal components. For example, a single non-linear FM signal component $z_1(t)$ can be written as:

$$z_1(t) = z_{1,a}(t) + z_{1,b}(t), \quad (21)$$

where $z_{1,a}(t)$ and $z_{1,b}(t)$ represent respectively the first and the second pieces constituting the signal $z_1(t)$. Its WVD is given by:

$$W_{z_1}(t,f) = W_{z_{1,a}}(t,f) + W_{z_{1,b}}(t,f) + \underbrace{2\mathrm{Re}\left\{\int_{\mathbb{R}} z_{1,a}\left(t+\frac{\tau}{2}\right) z_{1,b}^*\left(t-\frac{\tau}{2}\right) e^{-2j\pi f\tau}\,\mathrm{d}\tau\right\}}_{W_{z_{1,a} z_{1,b}}(t,f)}. \quad (22)$$

Cross-terms are oscillatory in nature, with the same order of magnitude as that of auto-terms [36], and they lie half-way in-between two signal components in (t, f) plane as illustrated in Fig. 4(a). The direction of oscillation of outer-terms is orthogonal to the line joining the two components while the direction of oscillation of inner-terms is partly orthogonal to the IF of a non-linear FM signal ([33], p. 63).

The ambiguity function is often used for designing QTFDs because it is the 2D-FT of the WVD, which then allows filtering in the (t, f) domain by a multiplication in the ambiguity domain. Specifically, Eq. (6) and Eq. (10) show that TFDs can be expressed as the result of a double convolution of (t, f) kernels with the WVD in the (t, f) domain or as the 2D-FT of the product of the Doppler–lag kernel with the ambiguity function. This leads to the following simple procedure for designing the QTFDs.

In the ambiguity domain, the cross-terms seem to appear away from the origin (see Fig. 4(b)) and are separated by a distance that is approximately equal to the frequency spacing in between two signal components as illustrated by Fig. 4(b). On the other hand, the auto-terms pass through the origin and appear between two cross-terms in the ambiguity domain. We note that the energies of the auto-terms are mostly concentrated around the origin as illustrated in Fig. 4(b).

QTFDs can then be defined to exploit the geometrical properties of auto-terms and cross-terms in the ambiguity domain by designing kernels using the "almost low-pass" characteristics of auto-terms and the high-pass characteristics of cross-terms in the ambiguity domain, resulting in kernels that discriminate and separate them. The cross-terms can therefore be reduced by applying a suitable 2D filter in the ambiguity domain and then transforming the ambiguity function back to the (t, f) domain.

The exact specifications of a 2D filter depend on the nature of the signal being analyzed. For example, for Fig. 4(b), a 2D filter that is elongated along the lag axis with low-pass characteristics along the Doppler axis will lead to a higher-resolution TFD. The result is shown in Fig. 4(c); and for illustration of this principle, the next subsection presents simple examples of relevant separable kernels that lead to high-resolution QTFDs.

2.2.3. Principles of separable kernel TFDs design

To illustrate this principle in clear terms, let us consider two special cases: lag-independent TFDs and Doppler-independent TFDs. These two cases of separable kernel TFDs can be used in certain specific situations related to signal characteristics, resulting in easy designs and implementations.

a) Lag-independent (LI) TFDs: Such lag-independent TFDs only perform smoothing along the time axis and are therefore characterized by $G_2(f) = \delta(f)$ or $g_2(\tau) = 1$. This kind of kernel is suitable for signals whose IF is parallel to the time axis in the (t, f) domain. In the ambiguity domain, such signals have most of their auto-term energy concentrated along the lag axis as these signals appear as an impulse along the frequency axis in the (t, f) domain, and as a constant along the lag axis in the ambiguity domain. The MBD kernel represents such an example of LI-TFD. It is defined in the ambiguity domain by the following kernel [1]:

$$g(\nu,\tau) = G_1(\nu) = \frac{|\Gamma(\beta + j\pi\nu)|^2}{\Gamma^2(\beta)}, \quad |\nu| \le \frac{1}{2},\ 0 \le \beta \le 1 \quad (23)$$

where ν and β are bounded to ensure that the MBD is a low-pass filter. Some physiological signals like HRV and EEG seizure have negligible FM; the direction of oscillation of outer-terms for such signals is parallel to the time axis. The MBD, a lag-independent distribution, is naturally suited for the analysis of such signals as it only performs smoothing along the time axis, thus avoiding the blurring of auto-terms in all directions usually caused by a 2D smoothing. To illustrate, let us consider a multi-component signal composed of two tones and one LFM,

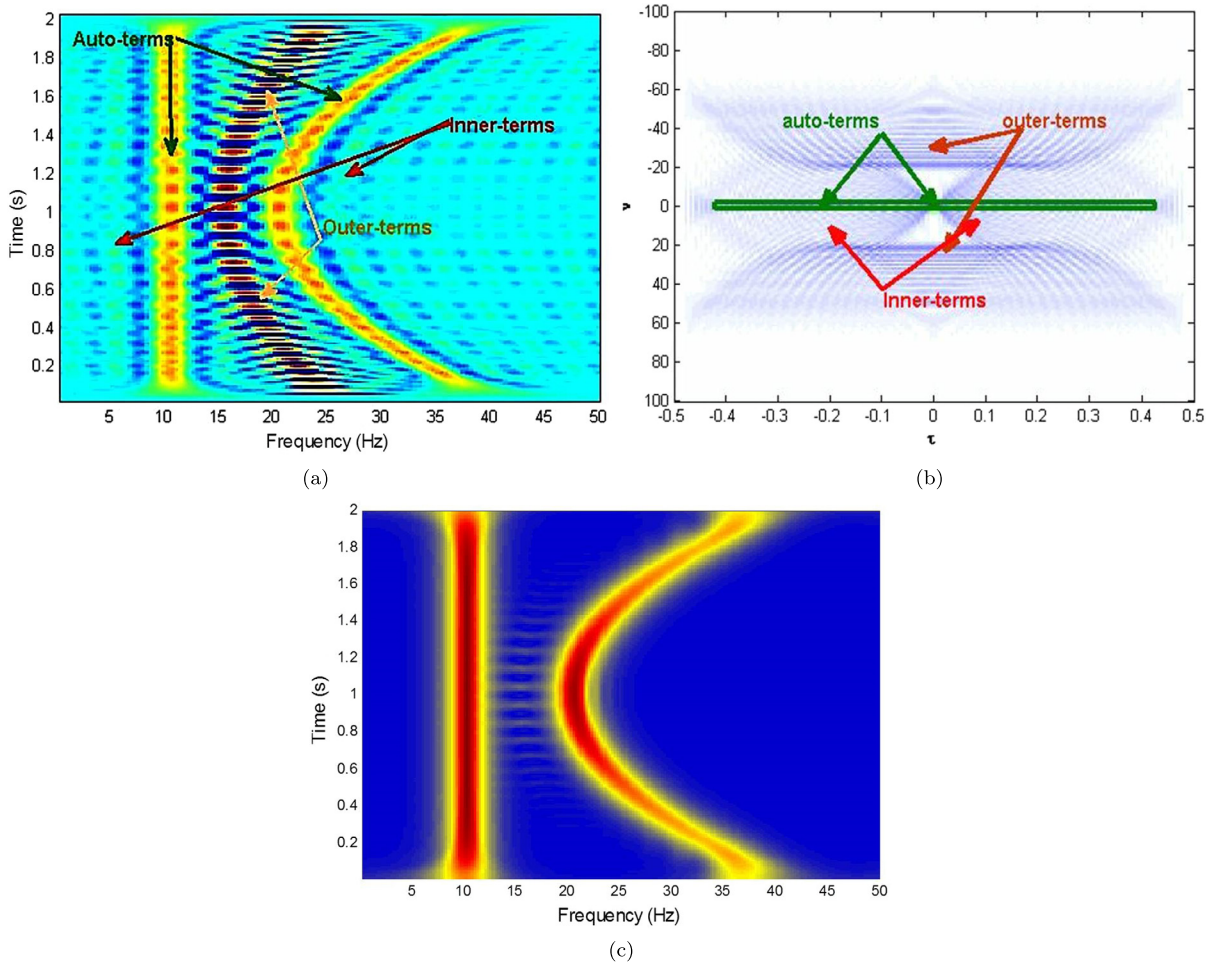


Fig. 4. Illustration of inner-term and outer-term artifacts using the sum of a quadratic FM signal and a linear FM signal. (a) (t, f) domain representation using the WVD. (b) Ambiguity domain representation. (c) (t, f) domain representation using the extended modified B distribution (EMBD).

and analyze it using the MBD, windowed WVD (discussed in the next subsection), and Hamming–Hanning kernel TFD [32] as shown in Fig. 5. The MBD, in Fig. 5(a), gives high energy concentration for the two tones but fails to give the same level of energy concentration for the LFM component as it performs smoothing only along the time axis. The Hamming–Hanning kernel TFD, shown in Fig. 5(b), gives a blurred representation for all the signal components due to its 2D smoothing kernel. The windowed WVD in Fig. 5(c) fails to suppress the cross-terms due to its frequency-only smoothing that is almost orthogonal to the direction of oscillation of cross-terms. Of course, for all the TFDs, further improvements can be made by optimizing window size and window shape following traditional DSP filter design criteria.

b) Doppler-independent (DI) TFDs: Such Doppler-independent TFDs only perform smoothing along the frequency axis and are characterized by $g_1(t) = \delta(t)$ or $G_1(\nu) = 1$. These DI types of kernels are appropriate for signals whose auto-terms are parallel to the frequency axis in the (t, f) domain. In the ambiguity domain, such signals have most of their auto-term energy concentrated along the Doppler axis. A simple example of a DI-TFD is the WVD windowed with a Hamming window along the lag axis expressed as:

$$g(\nu, \tau) = g_2(\tau) = 0.54 - 0.46\cos(2\pi\tau), \quad -0.5 \le \tau \le 0.5, \tag{24}$$

where τ is bounded to ensure that Hamming is a low-pass filter.

The DI-TFD is suited to real-world signals that have the characteristics described above, like for instance EEG spike signals which can be modeled by a train of impulses in the time-domain. The direction of oscillation of outer-terms for such signals is parallel to the frequency axis. Methods, such as the windowed WVD defined by Eq. (24), are more suitable for the analysis of this type of signals as they perform smoothing along the frequency axis only and avoid unnecessary blurring caused by a complete 2D filter. Fig. 6 shows the MBD, windowed WVD, and Hamming–Hanning kernel TFD of a signal composed of two impulses in the time domain and an LFM signal. The MBD, shown in Fig. 6(a), gives poor energy concentration for all the signal components as the direction of its smoothing kernel is orthogonal to the direction of energy concentration of the signal components. The Hamming–Hanning kernel TFD, shown in Fig. 6(b), performs 2D smoothing and therefore blurs the signal auto-terms, but its resolution is better than the MBD. The windowed WVD, shown in Fig. 6(c), results in a high-resolution (t, f) representation for the two impulses but it fails to give same resolution for the LFM component as it only performs smoothing along the frequency axis.

Apart from the LI and DI TFDs, most of the other separable kernel TFDs, like the B-distribution, the extended Modified B-distribution and the Compact support kernel employ smoothing filters along both time and frequency axes [33,32,10]. Some

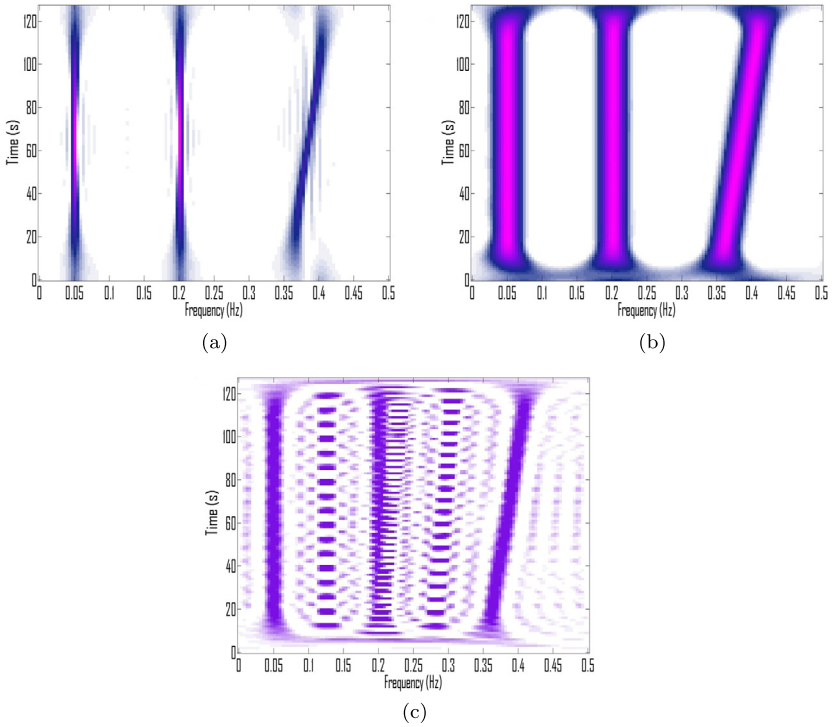


Fig. 5. (t, f) representations of a signal composed of two tones and one LFM. (a) The MBD ($\beta = 0.05$). (b) The Hamming–Hanning kernel TFD (window length = 32 s for both Doppler and lag axes). (c) The windowed WVD (rectangular window length = 31 s).

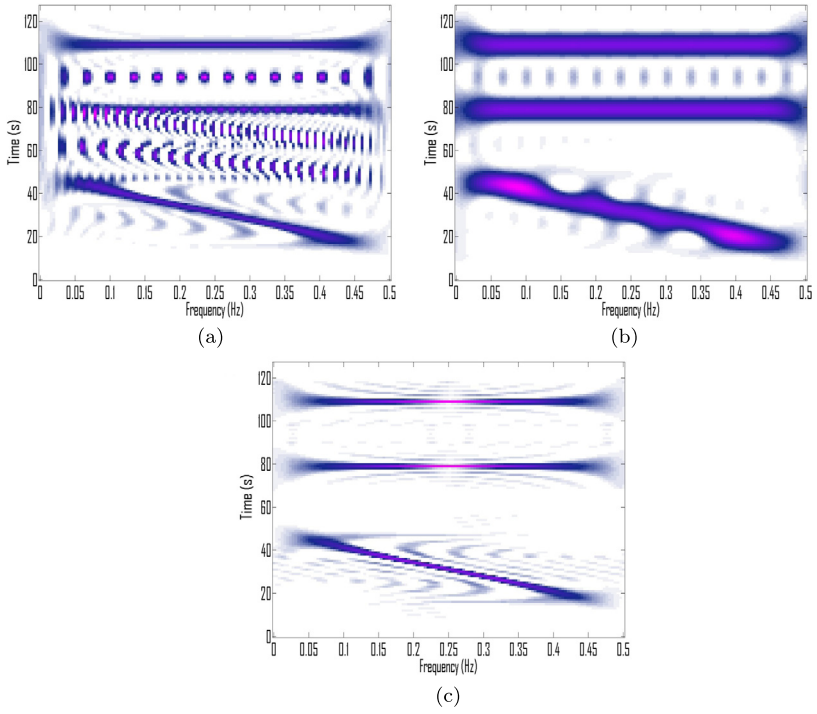


Fig. 6. TFD representation of a multi-component signal composed of two impulses and one LFM in the (t, f) domain using the (a) MBD ($\beta = 0.5$), (b) Hamming–Hanning-kernel TFD (window length = 32 s for both Doppler and lag axes), (c) windowed WVD (rectangular window length = 20 s).

examples of these particular kernels previously used in the literature or recently developed are given below.

2.2.4. Specific separable kernel TFDs

i. B-distribution Kernel ([33], p. 217): The B-distribution kernel is defined in the ambiguity domain as follows:

$$g(\nu,\tau) = g_2(\tau)G_1(\nu) = |\tau|^{\beta}\frac{|\Gamma(\beta + j\pi\nu)|^2}{2^{1-2\beta}\Gamma(2\beta)},$$
$$|\nu| \le 0.5,\ |\tau| \le 0.5,\ \text{and } 0 \le \beta \le 1, \tag{25}$$

where τ, ν and β are bounded to ensure that the B-distribution contains a low-pass filter.

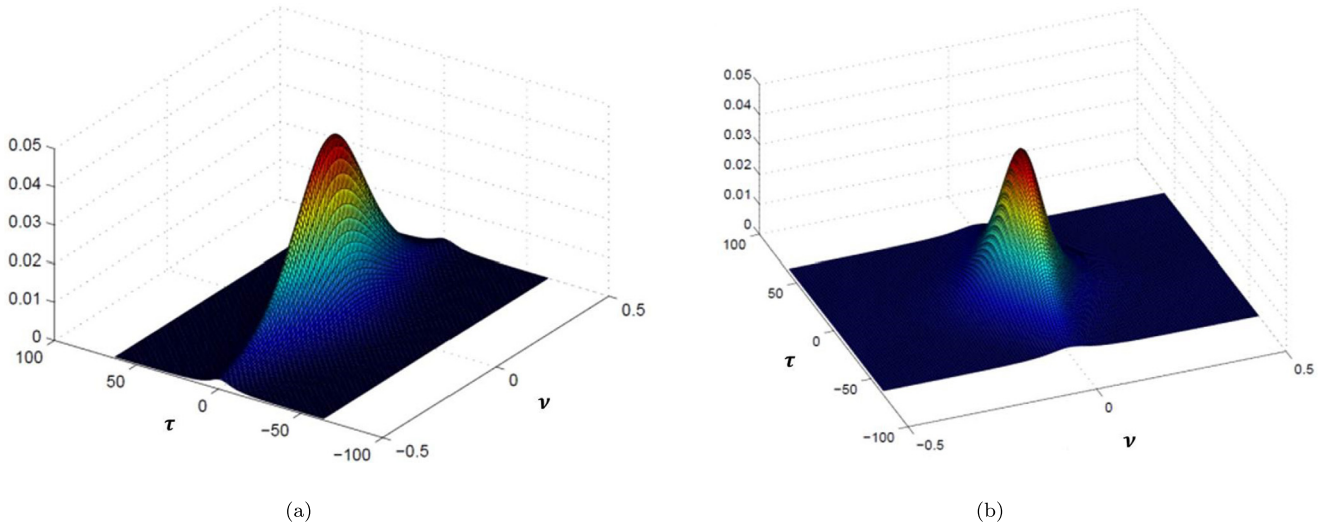


Fig. 7. (a) The EMBD kernel with dominant lag-direction. (b) The EMBD with dominant Doppler-direction.

The B-distribution kernel is the product of a low-pass filter $G_1(\nu) = \frac{|\Gamma(\beta+j\pi\nu)|^2}{2^{1-2\beta}\Gamma(2\beta)}$ and a high-pass filter $g_2(\tau) = |\tau|^\beta$. The B-distribution gives a high-resolution TFD for selected small values of β [37]. A disadvantage is that a zero at the origin appears as a consequence of the high pass filtering performed along the lag axis i.e. $g_2(0) = |0|^\beta = 0$. This results in a deterioration of the resolution of the B-distribution for certain types of signals [32]. This fact motivated its modification as detailed below.

ii. Modified B-distribution: As explained above, the MBD-kernel was designed to overcome some of the problems of the B-distribution, based on the following observations
   1 The B-distribution usually gives high resolution (t, f) representation for very small values of β $(0 < \beta \ll 1)$ ([33], p. 118).
   2 For small values of β, the lag window is approximately an all pass filter (except for $\tau = 0$), i.e., $g_2(\tau) \approx 1$.

   The kernel of the MBD includes only the Doppler window of the B-distribution kernels so as to avoid the distortion caused by high pass filtering along the lag-axis as shown in Eq. (25). The consequence is that the MBD gives the highest energy concentration only for signals whose IF does not vary rapidly with time, such as EEG seizure signals [38].

iii. Hamming–Hanning TFD: The corresponding Hamming–Hanning kernel is a 2D low-pass filter expressed as [32]:

$$g(\nu,\tau) = \big(0.54 - 0.46\cos(2\pi\tau)\big)\big(0.5 - 0.5\cos(\pi\nu)\big),$$
$$-\frac{1}{2} \le \nu \le \frac{1}{2}, \quad -\frac{1}{2} \le \tau \le \frac{1}{2}. \tag{26}$$

   This kernel results in a high-resolution TFD obtained by choosing a 2D low-pass filter along both lag and Doppler axes. This example is provided merely to show that any of the classical windows used for spectral analysis or digital filter design can be used as kernel filters, with specifications provided in the ambiguity domain.

iv. Extended modified B-distribution: As discussed earlier, previous studies have reported that the kernel of the lag-independent MBD mentioned in Eq. (25) gives optimal energy concentration for signals with negligible FM ([33], pp. 213–222), i.e., for signals whose IF is almost parallel to the time axis. The drawback of the lag-independent formulation of the MBD is that it does not allow smoothing along the frequency axis, thus limiting its scope for the analysis of signals whose auto-terms are parallel to the time axis. For example the MBD cannot be used for the analysis of a signal that is composed of a train of impulses as observed in some newborn EEG signals, as it then would give poor results [1]. The EMBD exhibits improvement in some situations as it extends the MBD by applying its kernel filter along both lag and Doppler axes, resulting in the expression [32]:

$$g(\nu,\tau) = \frac{|\Gamma(\beta+j\pi\nu)|^2}{\Gamma^2(\beta)}\,\frac{|\Gamma(\alpha+j\pi\tau)|^2}{\Gamma^2(\alpha)}, \tag{27}$$

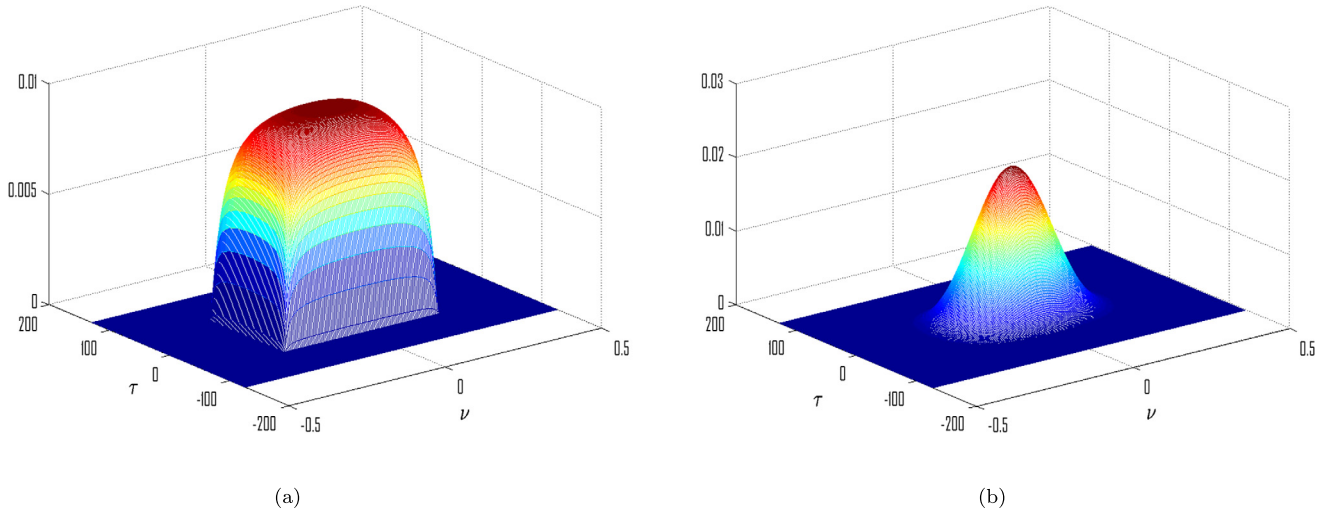
   where $-0.5 \le \nu \le 0.5$, $-0.5 \le \tau \le 0.5$, $0 \le \beta \le 1$ and $0 \le \alpha \le 1$. The lengths of the Doppler and lag windows are controlled by separate parameters α and β respectively. The extra degree of freedom in the formulation of the EMBD allows to independently adjust the lengths of the windows along both lag and Doppler axes as illustrated in Fig. 7. This makes it a more useful tool for the analysis of real-life signals, such as fetal movement signals, EEG signals with seizure and EEG spike signals. However, the EMBD has only one parameter for each kernel filter or window (that is α for lag and β for Doppler) to control both shape and size. Thus the EMBD does not allow adapting both length and shape of the smoothing window of the kernel filter independently, although it is an improvement on both the BD and the MBD.

v. Compact support kernel TFD: Such compact support kernels (CSK) are designed to vanish outside a given range in the ambiguity domain. Unlike Gaussian windows they do not have infinite length, so there is no need to truncate those using rectangular windows that may cause loss of information. These TFDs have been shown to outperform other kernel-based methods in terms of their ability to suppress cross-terms while retaining the resolution of auto-terms in some cases [10]. Such high-resolution performance is achieved by these kernels by combining their compact support with a flexibility to adjust both shape and size independently, as discussed below.

   In order to explain the characteristics of the CSK, let us first observe in the formulation below that its kernel includes two components [10]:

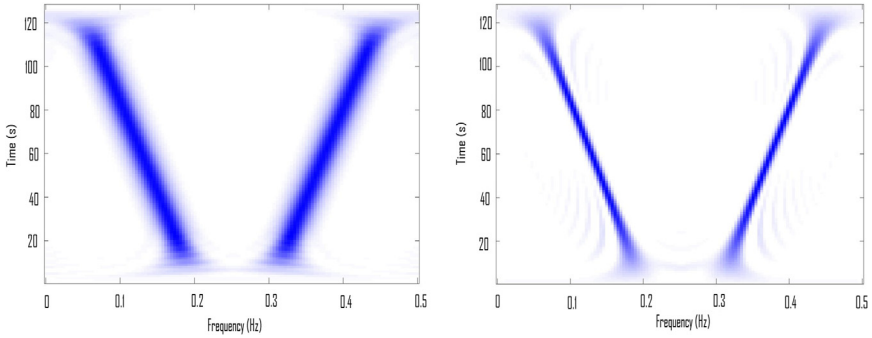
$$g(\nu,\tau) = \begin{cases} e^{2c}e^{\frac{cD^2}{\nu^2-D^2}+\frac{cD^2}{\tau^2-D^2}}, & \begin{cases}\nu^2 < D^2\\ \tau^2 < D^2\end{cases} \\ 0, & \text{otherwise} \end{cases} \tag{28}$$

   In the above, the two branches formed by the Doppler window and lag window of the CSK are respectively given by



(a) (b)

Fig. 8. (a) The CSK as a Quasi-Rectangular window. (b) The CSK as a quasi-Gaussian window.



(a) (b)

Fig. 9. (a) The Spectrogram (Bartlet window length = 55 s). (b) Hamming–Hanning kernel TFD (lag window length = 128 s, Doppler window length = 0.07 Hz).

$$G_1(\nu) = \begin{cases} e^c e^{\frac{cD^2}{\nu^2 - D^2}}, & |\nu| < D \\ 0, & \text{otherwise} \end{cases} \quad \text{and}$$

$$g_2(\tau) = \begin{cases} e^c e^{\frac{cD^2}{\tau^2 - D^2}}, & |\tau| < D \\ 0, & \text{otherwise} \end{cases} \tag{29}$$

The above equations show that the shape and size of both Doppler and lag windows are determined respectively by the parameters c and D. These parameters allow the CSK to adapt both kernel length and shape. For example, the shape of the CSK in the ambiguity domain can be adjusted from the quasi-rectangular window to the quasi-Gaussian window as shown in Fig. 8. However, the CSK definition restricts the Doppler and lag windows to be of same length, so that the smoothing along time or frequency axis cannot be adjusted independently. This limitation is overcome by a new design procedure described in Section 2.3.

2.2.5. Performance comparison: separable kernel TFD vs spectrogram

To illustrate the potential advantage of using separable kernel TFDs over the Spectrogram in terms of energy concentration and resolution properties, let us consider a two component signal defined as:

$$s(t) = \begin{cases} \cos(2\pi(0.2t - 0.0006t^2)) + \cos(2\pi(0.3t + 0.0006t^2)), \\ \quad 0 \le t \le 127, \\ 0, \quad \text{otherwise}, \end{cases} \tag{30}$$

The Spectrogram and Hamming–Hanning kernel based TFD for this signal are shown in Fig. 9. The parameters of both TFDs were optimized based on visual analysis.

Fig. 9 shows that the separable kernel TFD gives high energy concentration for the given signal as compared to the Spectrogram. This improved performance of the Hamming–Hanning kernel based TFD is due to additional flexibility in its formulation to independently adjust smoothing along time and frequency axes.

2.3. An advanced TFD design with improved resolution: the CKD

2.3.1. Non-uniform compact support kernel

Eq. (29) indicates that neither the shape nor the size of the Doppler and lag windows can be adjusted independently of each other. This shortcoming limits the application of the CSK TFD to the analysis of signals whose energy is homogeneously distributed in the ambiguity domain. As a consequence, the CSK TFD cannot deal optimally with signals like EEG seizure signals, whose IF is almost parallel to the time axis, or EEG spike signals, whose IF is almost parallel to the frequency axis.

To illustrate the limitations of the CSK TFD, let us consider a multi-component signal composed of two LFM signals. Fig. 10(a) shows the ambiguity function modulus of this signal and a CSK with optimal parameters. The CSK, even with optimal parameters, fails to retain all the auto-term energy in the ambiguity domain, thus resulting in a blurred TFD as shown in Fig. 10(b). Fig. 10(c) illustrates the ambiguity domain filtered by a separable kernel that is elongated along the lag axis to retain all auto-term energy. The separable kernel filters cross-terms while retaining all the auto-

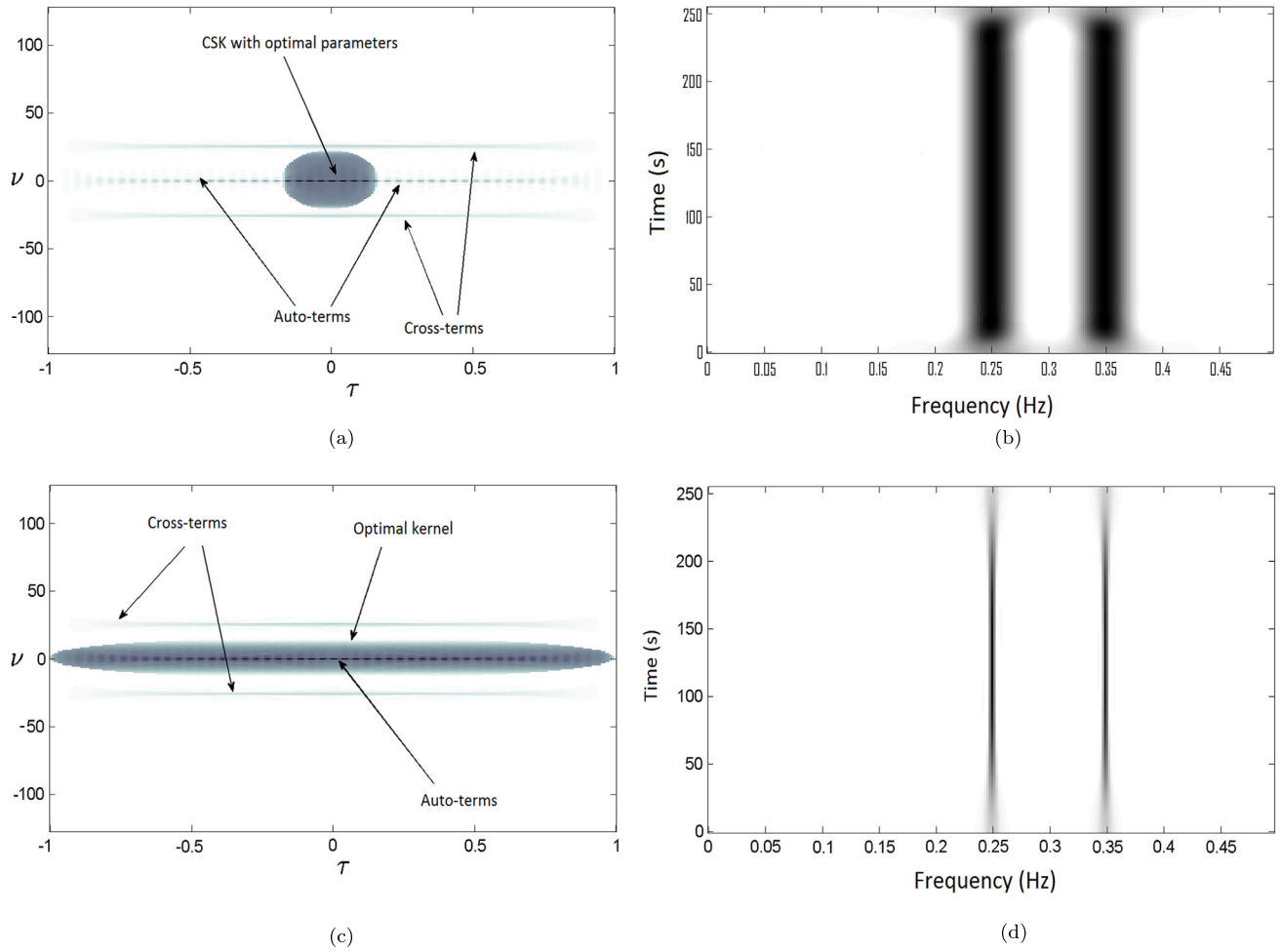


Fig. 10. Ambiguity functions and TFDs of two close tones using a CSK TFD and a separable kernel TFD. (a) The optimized CSK in ambiguity domain. (b) CSK-TFD with optimized parameters. (c) Ambiguity domain representation of the optimal kernel TFD. (d) TFD representation after the application of the optimal kernel.

terms, resulting in a higher-resolution TFD as shown in Fig. 10(d). The low-resolution performance of the CSK is due to the restriction in its formulation to have equal length for both Doppler and lag windows.

2.3.2. Derivation of the extended compact support kernel

The CSK can be extended by simply modifying the formulation of Doppler and lag windows so that their lengths can be adjusted independently. Such modified Doppler and lag windows can then be expressed as:

$$G_1(\nu) = \begin{cases} e^c e^{\frac{cD^2}{\nu^2 - D^2}}, & |\nu| < D \\ 0, & \text{otherwise} \end{cases} \quad \text{and}$$

$$g_2(\tau) = \begin{cases} e^c e^{\frac{cE^2}{\tau^2 - E^2}}, & |\tau| < E, \\ 0, & \text{otherwise.} \end{cases} \tag{31}$$

As mentioned above, the lengths of the modified Doppler and lag windows now depend on two parameters D and E. These parameters are selected on the basis of shape and orientation of auto-terms in the ambiguity domain so that the maximum energy of auto-terms is retained and cross-terms are reduced. This leads to the following formulation of the Extended Compact Kernel (ECK):

$$g(\nu, \tau) = G_1(\nu) g_2(\tau) = \begin{cases} e^{2c} e^{\frac{cD^2}{\nu^2 - D^2} + \frac{cE^2}{\tau^2 - E^2}}, & |\nu| < D, \ |\tau| < E, \\ 0, & \text{otherwise.} \end{cases} \tag{32}$$

The additional degree of freedom in the formulation of the ECK to independently adjust the lengths of the Doppler and lag windows is illustrated in Fig. 11.

The resulting ECK performs smoothing along both time and frequency axes. Note that, like many other TFDs, it does not rigorously fulfill mathematical properties like time support, frequency support, time and frequency marginals, and positivity ([33], p. 216). However, it satisfies the mathematical properties of realness, energy conservation and (t, f) shift invariance [10], which are sufficient and convenient for a number of applications such as classification, as justified below.

1. Real TFDs decrease the dimensionality of the classification problem as compared to complex TFDs.
2. Time-shift and frequency-shift invariance are needed to ensure the consistent performance of pattern classification algorithm.
3. The conservation of the total energy is necessary to preserve the relative energy between features of different classes of data.

The TFD defined by the ECK kernel may be referred to as ECK TFD, but this abbreviation is further shortened to Compact Kernel Distribution (CKD) for simplicity.

2.3.3. Discrete-time implementation of the CKD

The time–lag domain is often used for the discrete-time implementation of TFDs for computational efficiency ([33], pp. 268–278). The discrete-time time–lag formulation of the TFD of the analytic

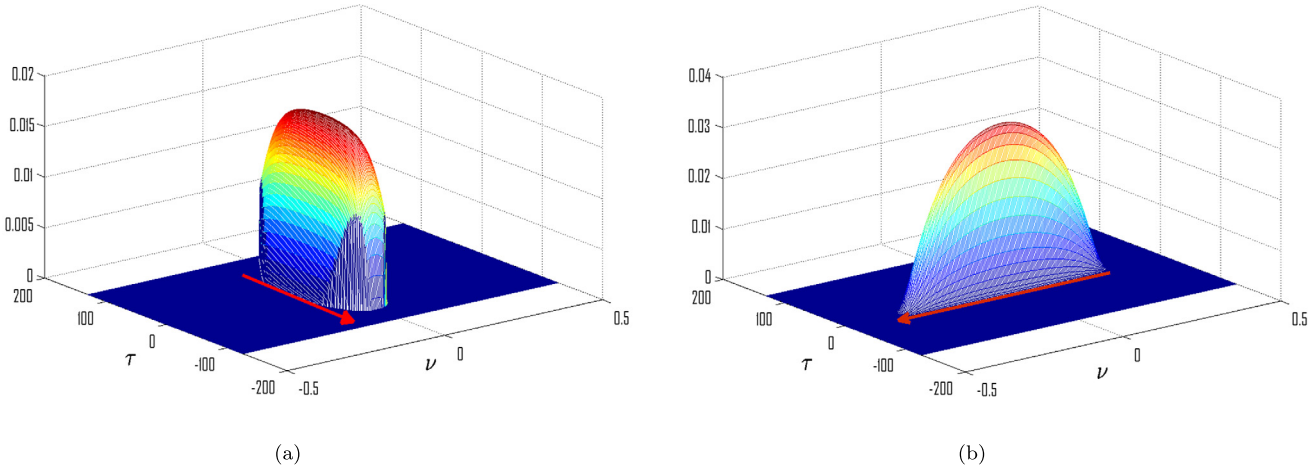


Fig. 11. (a) The ECK: Quasi-similar to the rectangular window with a dominant Doppler direction. (b) The ECK: Quasi-similar to the Gaussian window with a dominant lag direction. (The dominant direction is shown by the red arrow). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

associate $z[n]$ of the N-point real signal $s[n]$ can be expressed as ([33], pp. 237):

$$\rho[n,k] = 2\underset{m\to k}{\mathrm{DFT}}\big(G[n,m]\underset{n}{*}\big(z[n+m]z^*[n-m]\big)\big). \tag{33}$$

The discrete-time variable n and discrete-frequency variable k are related to the continuous-time and continuous-frequency variables by the following expressions ([33], pp. 232–241)

$$n = tf_s \quad \text{and} \quad k = f\frac{2M}{f_s}, \tag{34}$$

where f_s is the sampling frequency. The resulting discrete TFD is a matrix of size $N \times M$, where N is the number of time-domain signal samples, and M is the number of frequency-domain samples. Following Eq. (32), the discrete time–lag formulation of the ECK is expressed as:

$$G[n,m] = \underset{l\to n}{\mathrm{DFT}}\big(e^{\frac{cD^2}{(\frac{l}{N})^2-D^2}}\big)e^{2c}e^{\frac{cE^2}{(\frac{m}{N})^2-E^2}}. \tag{35}$$

Its properties are discussed in the next section with respect to performance.

2.3.4. Performance evaluation of the CKD

Let us first consider an LFM signal as:

$$s(t) = A\,\mathrm{rect}\left[\frac{t}{T}\right]\cos\left(2\pi\left(f_0 t + \frac{\alpha}{2}t^2\right) + \psi\right), \tag{36}$$

where A is the amplitude (assumed to be equal to 1), T is the total duration and ψ represents a phase offset (ψ is often equal to 0), and $\mathrm{rect}[\frac{t}{T}]$ represents a rectangular window defined by:

$$\mathrm{rect}[\tau] = \begin{cases} 1 & -\frac{1}{2} \le \tau \le \frac{1}{2} \\ 0, & \text{otherwise.} \end{cases} \tag{37}$$

The frequency rate α is estimated by the difference between the maximal frequency $f_{\max}$ and the initial frequency f_0, divided by the total duration T:

$$\alpha = \frac{f_{\max} - f_0}{T}. \tag{38}$$

The performance of the CKD defined above is evaluated using the following signals:

a) A synthetic multi-component signal composed of two LFM components defined as:

$$s(t) = \mathrm{rect}\left[\frac{t-128}{T}\right]\left(\cos\left(2\pi\left(0.1t + \frac{0.0006}{2}t^2\right)\right) + \cos\left(2\pi\left(0.15t + \frac{0.0006}{2}t^2\right)\right)\right). \tag{39}$$

The sampling frequency is 1 Hz, the signal duration is equal to $T = 255$ s and the frequency rate α for both signals is the same and is given by: $\alpha = \frac{f_{\max,1} - f_{0,1}}{T} = \frac{(0.253-0.1)}{255} = 0.0006$, where $f_{\max,1}$ and $f_{0,1}$ are respectively the maximal frequency and the initial frequency of the first LFM signal.

b) An EEG seizure signal of 8-s duration sampled at 32 Hz.

c) A fetal movement signal of 2.5-s duration sampled at 100 Hz.

These signals are analyzed using the TFDs defined in previous sections, including the EMBD, S-method, CSK-TFD, MBD, WVD, and CKD. The results are shown in Figs. 12, 13 and 14. The parameters of these TFDs for the synthetic signals are optimized with respect to Boashash–Sucic's criterion [38] while for the real-life signal, their parameters are optimized on the basis of visual inspection (see Appendix A for the definition of this resolution measure).

The resulting 'P' values for the TFDs of the synthetic signal are shown in Table 1. On the basis of both visual and quantitative analysis, the CKD outperforms all the other TFDs for all these signals due to a better ability to reduce cross-terms while maintaining a high resolution of auto-terms. The S-method is the second best performing TFD for synthetic, real-life EEG and fetal movement signals. Note that the performance of the MBD can be significantly improved by using a rectangular window of length smaller than the signal length, but such modification will make the MBD a lag dependent distribution, i.e., an EMBD.

The CKD gives the best performance for signals whose auto-terms are nearly parallel to either time or frequency axis. However, its performance degrades for signals whose auto-terms have a specific direction away from the time axis or frequency axis in the (t, f) domain. In such cases, the optimum can be reached using directional filtering as a post-processing operation to design a high-resolution TFD as detailed in Section 4.2.

3. Performance assessment for multi-component instantaneous frequency estimation

For stationary signals, frequency estimation is used in a number of applications including music and speech to estimate the relative

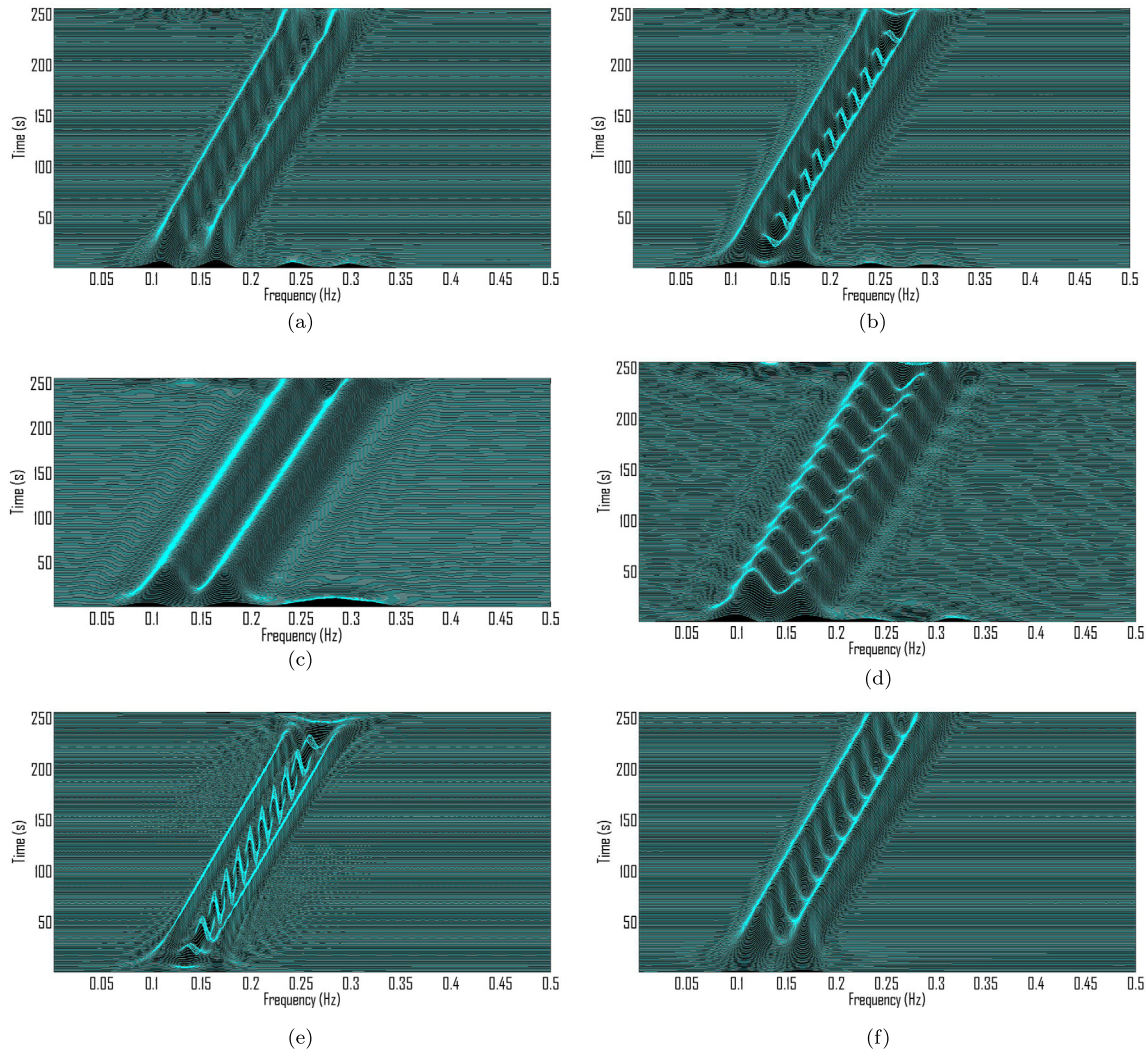


Fig. 12. The TFDs of a multi-component signal composed of two LFM signals. (a) The CKD ($c = 1$, $D = 0.05$, $E = 0.12$). (b) The EMBD ($\alpha = 0.18$, $\beta = 0.06$). (c) The CSK-TFD ($c = 0.6$, $D = 0.05$). (d) The WVD. (e) The MBD ($\beta = 0.16$). (f) The S-method (Hamming window length = 4 s).

strengths of each frequency component [39]. For non-stationary signal processing, the method extends to IF estimation which represents the variation of the dominant frequency as a function of time [17]. This concept of IF is intrinsically linked to the concept of TFDs as the IF estimation capability of TFDs is often used as a criterion to evaluate their performance, including robustness against noise. The performance of the previously defined TFDs is comparatively assessed below for both simulated and real-world signals in terms of IF performance estimation. This requires first to define clearly what is being estimated and the criteria for comparison. Let us consider first the simple case of mono component signal IF estimation.

3.1. Meaning and estimation of the instantaneous frequency for mono-component signals

In TFD based IF estimation the criteria of goodness for mono component signals are: no bias and minimum variance. These IF estimates can be defined as the peaks of TFDs or as their first moments. One aim of defining high-resolution TFDs is to also yield high concentration TFDs, resulting in efficient IF estimates.

As the concept of IF was introduced to represent the spectral variations of non-stationary signals [17], a simple example is the FM signals, for which the IF is the FM law. The easiest is the LFM, which is used in various applications, such as "vertical seismic prospecting" (VSP), where it is desired to study parameters such as soil absorption and dispersion [4]. For such signals, the WVD is the optimal representation in terms of energy concentration.

Let us consider a mono-component analytic signal $z(t)$ modeled as $z(t) = a(t)e^{j\phi(t)}$, such that $a(t)$ is a low-frequency signal whose spectrum does not overlap with the high-frequency signal $e^{j\phi(t)}$ as per Eq. (1) and $\phi(t)$ is defined by $\phi(t) = \arctan(\mathrm{Im}(z(t))/\mathrm{Re}(z(t))$. The IF of the mono-component signal is defined as the time derivative of its phase:

$$f(t) = \frac{1}{2\pi}\frac{d\phi(t)}{dt}. \tag{40}$$

A comparative review of fundamental algorithms for IF estimation appeared in [40]. More recent advances are reviewed next, setting the stage for the more difficult case in next section that is used as a test for comparing the new TFDs defined in the previous section.

3.2. IF estimation for mono-component signals

The main (t, f) approach to IF estimation follows the simple principle that the IF of a mono-component signal at each time instant can be estimated by detecting the location of peaks along the frequency axis in the (t, f) plane.

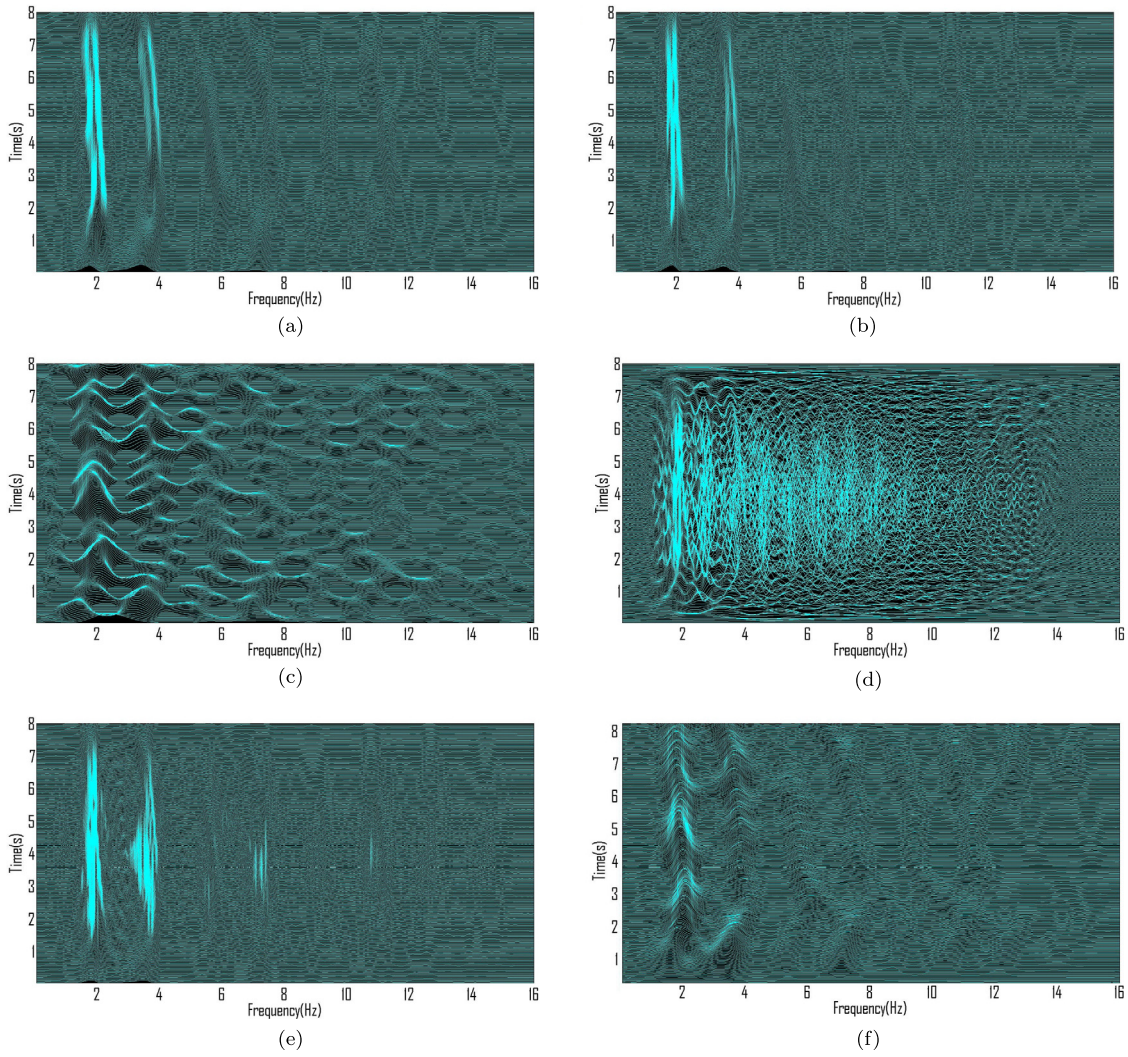


Fig. 13. TFDs of a seizure signal. (a) The CKD ($c = 1$, $D = 0.2$, $E = 0.05$). (b) The EMBD ($\alpha = 0.02$, $\beta = 0.75$). (c) The CSK-TFD ($c = 0.1$, $D = 0.06$). (d) The WVD. (e) The MBD ($\beta = 0.03$). (f) The S-method (Hamming window length = 4 s).

Table 1
Resolution performance measure P of selected TFDs for a multi-component signal composed of two parallel LFM signals (The optimized parameters are obtained using an exhaustive search of the parameter that maximizes the resolution measure; it is not "optimal" in the classical sense. See details in Appendix A).

Time frequency distribution	Optimized parameters	P
Compact Kernel Distribution	$C = 1$, $D = 0.05$, $E = 0.125$	0.8919
Extended Modified B distribution	$\alpha = 0.18$, $\beta = 0.06$	0.8376
Compact Support Kernel	$C = 0.6$, $D = 0.05$	0.7854
Wigner–Ville Distribution	Not available	0.7063
Modified B-distribution	$\beta = 0.16$	0.7628
S-Method	Hamming window length = 4 s	0.8778

Several algorithms are based on the above principle. For example, the WVD gives an ideal estimate for mono-component LFM signals so it can be used as an IF estimator in such cases. However, its estimate is biased for non-linearly FM signals [41], and so other methods are needed. In the general case, the bias can be overcome by estimating the IF from a windowed WVD, but windowing increases the variance of the IF estimate. An optimal window length for an optimum bias-variance tradeoff at each time instant can be estimated adaptively [41]. Following this approach, multiple WVDs can be used to obtain multiple estimates of the IF as well as the confidence interval for each estimate. Starting from the smallest window, the confidence intervals of the IF estimates of successive windows are tested for overlap till the intersection of the confidence intervals of two successive windows results in a null set. The IF estimate of the longest window fulfilling the criterion of overlapping of confidence intervals is then selected as an optimal estimate. This method is known as the ICI rule where ICI means intersection of confidence intervals [41]. A modified approach further improves the accuracy of the ICI algorithm by taking into account the amount of overlap between two successive confidence intervals. Other iterative IF estimation algorithms can obtain the initial estimate of the IF from the Spectrogram [42]. The IP, defined as the integration of the IF, is then used to demodulate the original signal. The demodulation shifts the spectrum of the original signal around the zero frequency. The IF of the demodulated signal is then estimated again and this process is iterated till there is no further change in two consecutive estimates of the IF.

Most QTFDs fail to accurately estimate the IF at very low SNR. In such scenarios improvements can be obtained by taking into account the main direction in the (t, f) plane where there is most energy concentration. For example, the adaptive combination of the directionally smoothed WVDs was shown to obtain a (t, f) representation that has high concentration of signal energy along the IF of the signal even at a very low SNR [43]. Such methods need to be further extended to deal with the more general multi-component case as addressed in the next section.

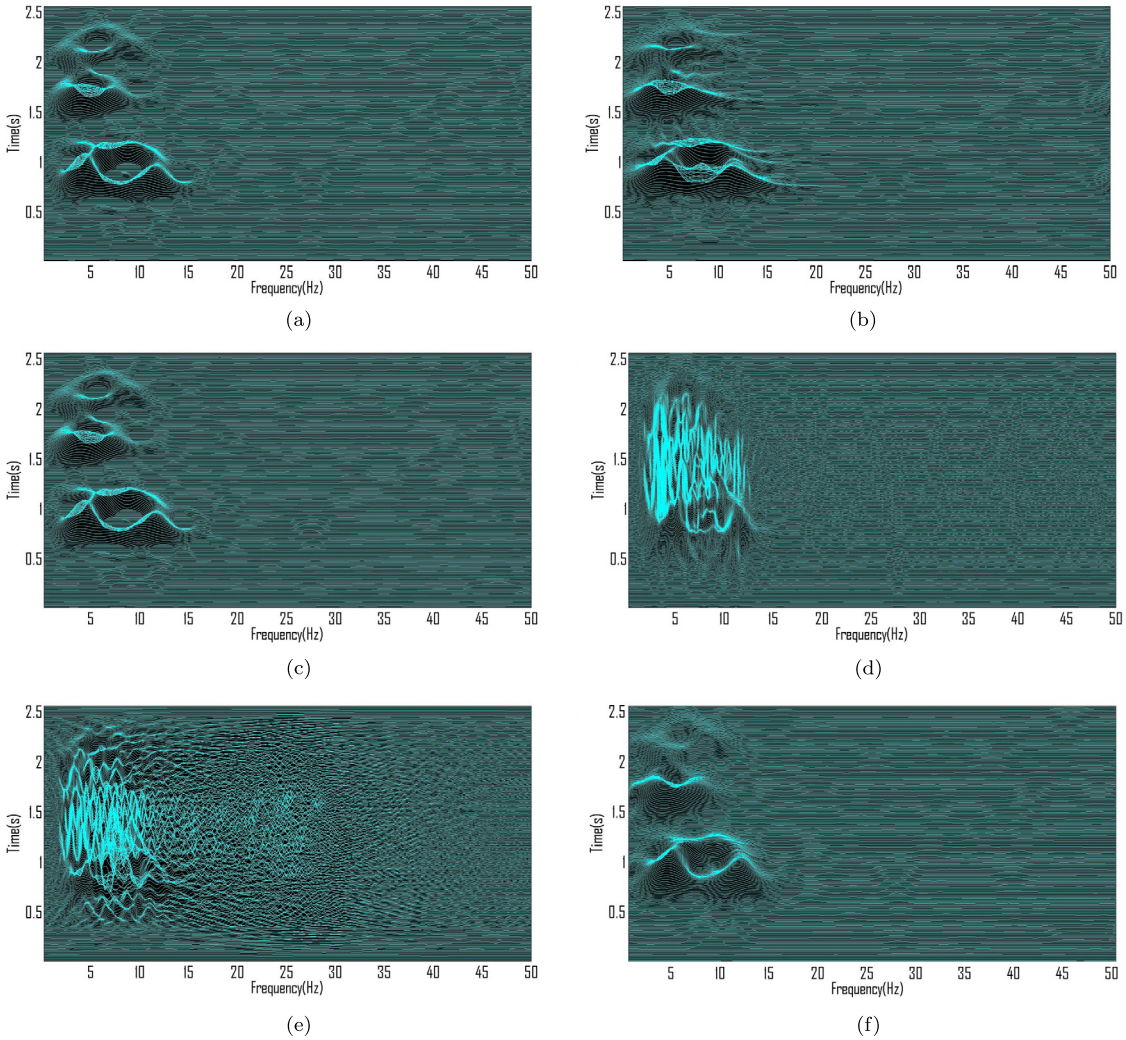


Fig. 14. TFDs of a fetal movement signal. (a) The CKD ($c = 1$, $D = 0.075$, $E = 0.075$). (b) The EMBD ($\alpha = 0.125$, $\beta = 0.075$). (c) The CSK-TFD ($c = 1$, $D = 0.075$). (d) The WVD. (e) The MBD ($\beta = 0.05$). (f) The S-method (Hamming window length = 4 s).

3.3. IF estimation for multi-component signals

For a multi-component signal, the (t, f) approach follows the principle that a multi-component analytic signal may be expressed as the sum of N_c mono-component signals, i.e.,

$$z(t) = \sum_{i=1}^{N_c} z_i(t) = \sum_{i=1}^{N_c} a_i(t) e^{j\phi_i(t)} \tag{41}$$

The IF of the ith signal component is defined by the time derivative of its phase as

$$f_i(t) = \frac{1}{2\pi} \frac{d\phi_i(t)}{dt} \tag{42}$$

If the individual components of the signal are FM signals, then the multi-component signal appears as multiple ridges in the (t, f) plane. Multi-component IF estimation algorithms need then to involve both detection and linking of related peaks to estimate the IF of each signal component. The linkage stage is needed so that the IF of each component can be tracked accurately and without mix-up with the IFs of other components.

3.3.1. IF linking algorithms

Image processing techniques can be used to link all the points of each IF in the (t, f) domain [44]. The key stages involve:

1. Detecting peaks in a TFD by using first and second order partial derivatives along the frequency axis.
2. Linking the detected peaks using some neighborhood-connectivity criteria [44]. For example, using a 10-neighborhood connectivity criterion, a peak belonging to the IF of a signal component must have at least one other peak of the IF of the same signal component in its 10-neighborhood and it should not have any peak of the IF of any other signal component in this selected neighborhood. The 10-neighborhood of a peak at the location (x, y) is defined as a set of all pixels at the following locations [44].

$$\begin{Bmatrix} (x+1,y) & (x+1,y+1) & (x+1,y+2) & (x+1,y-1) & (x+1,y-2) \\ (x-1,y) & (x-1,y+1) & (x-1,y-2) & (x-1,y-1) & (x-1,y-2) \end{Bmatrix} \tag{43}$$

This image processing linking algorithm is fully automatic and does not require prior information like the number of components, their IF laws and SNR.

3.3.2. Multi-component IF estimation based on component extraction and ICI algorithm

Let us consider for illustration another method, the ICI based IF estimation algorithm which was extended to multi-component signals by first extracting signal components using a blind source separation algorithm [45]. This algorithm iteratively extracts signal

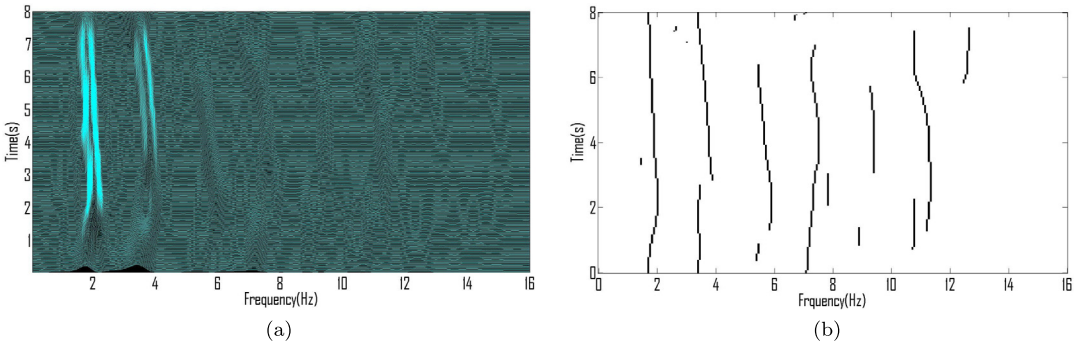


Fig. 15. (a) The CKD ($c = 1$, $D = 0.2$, $E = 0.05$) of an EEG signal. (b) The IFs of signal components estimated using the method described in [44].

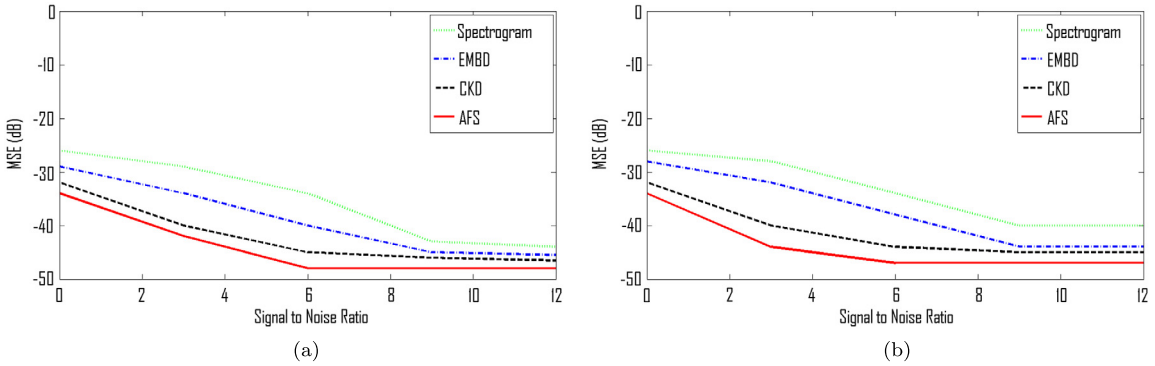


Fig. 16. (a) The MSE of the IF estimates of $s_1(t)$ obtained using high-resolution TFDs vs. SNR. (b) The MSE of the IF estimates of $s_2(t)$ obtained using high-resolution TFDs vs. SNR. In all these cases the IFs are estimated using the component extraction procedure [47].

components from the (t, f) plane till the total energy of the signal falls below a certain threshold. The disadvantage of this algorithm is its dependence on prior information like total number of signal components or SNR of a signal.

A comparative study has shown that the above image processing algorithm based on connected component linking is more suitable for estimating the IF of the physiological signals considered as it is computationally efficient as compared to the component extraction method [46]. Fig. 15 illustrates an example of IF estimation of a multi-component EEG signal using the image processing technique described in [44]. Both ICI with blind source separation and connected component linking algorithms can only be applied to (t, f) separable signals, i.e., signals whose components do not overlap with each other in the (t, f) domain. Many of the real-world signals do not fulfill this property so these methods are not general and should be applied only to relevant classes of signals, therefore requiring a-priori knowledge.

In order to design high performance multi-component IF estimation algorithms for the general case, it is important to remember that the performance of multi-component IF estimation algorithms depend on the resolution and cross-term suppression capabilities of the TFD employed. Recent studies have indicated that data-dependent high-resolution TFDs like the Adaptive Fractional Spectrogram (AFS) [47] and variable-bandwidth filter based heterogeneous TFD [9] can outperform other TFDs in terms of their ability to accurately estimate the IFs of closely spaced signal components at low SNR. The improvement is naturally obtained using an adaptive procedure so as to get a "perfect match" with the data. The previous TFDs are therefore compared below in terms of their IF estimation performance. The AFS is included in the comparison as an example of data dependent TFD.

3.3.3. Comparison of TFDs in terms of their multi-component IF estimation performance

In order to evaluate the IF estimation capability of the TFDs designed above, let us consider a two-component real signal expressed as:

$$s(t) = s_1(t) + s_2(t), \tag{44}$$

where

$$s_1(t) = \text{rect}\left[\frac{t-128}{T}\right]\cos\left(2\pi\left(0.1t + \frac{0.15}{T}t^2\right)\right), \tag{45}$$

$$s_2(t) = \text{rect}\left[\frac{t-128}{T}\right]\cos\left(2\pi\left(0.15t + \frac{0.15}{T}t^2\right)\right), \tag{46}$$

with time duration $T = 256$ s and sampling frequency $f_s = 1$ Hz. The corresponding IFs are

$$f_1(t) = 0.3\frac{t}{T} + 0.1, \tag{47}$$

$$f_2(t) = 0.3\frac{t}{T} + 0.15. \tag{48}$$

The TFDs considered are the EMBD, Spectrogram, CKD and AFS; for the latter the IF is estimated using the component extraction procedure as discussed in [47]. The mean square error (MSE) is estimated by performing 200 Monte Carlo simulations [48]. Fig. 16 displays the results of the IF estimates. The results indicate that the CKD gives more accurate estimates of the IFs as compared to other non-adaptive TFDs, and naturally, the AFS does even better due to the adaptation.

The above findings indicate that an accurate estimation of the IF of the signal components is directly related to the cross-term suppression and auto-term resolution properties of (t, f) methods employed. Therefore, there is a need to develop better methods

for improving the resolution of TFDs. The following section will discuss image processing methods for improving the resolution of TFDs.

4. Time–frequency image processing techniques for TFD enhancement

Clarity of representation is essential for TFDs, so that their relevant features can be read, selected and extracted. As all QTFDs suffer from limitations such as inherent compromise between cross-term suppression and auto-term resolution, it may adversely affect the extraction of some relevant features from (t, f) images as well as components separation. To improve the readability, let us consider the TFD as a (t, f) image. Then, image processing techniques can be used as a processing stage to either improve the resolution of cross-term-free low-resolution TFDs (e.g., the Spectrogram) using de-blurring methods or further suppress cross-terms in high-resolution TFDs (e.g., the WVD). Section 4.1 reviews existing image processing techniques for TFD image enhancement, and Section 4.2 presents an improved novel image processing technique for improving the resolution of the class of QTFDs considered and therefore allowing improved multi-component IF estimation.

4.1. TFD enhancement using image processing techniques

Given the number of different ways to generate a high-resolution TFD and the imperfections associated with each one of them, a natural approach is to process the TFD image using image processing techniques to improve the clarity of the image. Three types of processing are reviewed in the following sections: (t, f) image de-blurring, (t, f) de-noising and (t, f) image artifact suppression.

4.1.1. Time–frequency image de-blurring

Image de-convolution techniques are used for the estimation of high resolution images from blurred ones. For example, these techniques have been applied to estimate an original high-resolution TFD from the Spectrogram, which can be interpreted as a blurred version of the true (t, f) image [34]. Another method reduces the blurring in the Spectrogram by using artificial neural networks [49]. This technique estimates the inverse of the blurring function using neural networks, which are trained using a pre-computed Spectrogram and the WVD. The trained neural networks are then applied to the Spectrogram to estimate the true high-resolution TFD. The disadvantage of this technique is that it results in a discontinuous TFD [49].

4.1.2. Time–frequency image de-noising

One way to enhance (t, f) images is to filter out noise to improve their SNR, using image de-noising techniques. Such techniques often assume an additive white noise model to account for the distribution of noise in images. Unfortunately, this model is not valid in general for TFDs as earlier studies have shown, that apart from the WVD, noise is not uniformly distributed in (t, f) images [50]. The influence of noise is stronger in the region of support of auto-terms as noise appearing outside the region of support of the auto-terms is reduced due to the low-pass filtering applied in the ambiguity domain. So, (t, f) image de-noising algorithms should be designed by taking into account the characteristics of noise distribution in QTFDs.

Some of the image de-noising algorithms that have been applied to (t, f) images are discussed below.

a) **Singular Value Decomposition (SVD) based methods:** As signals represented in the (t, f) domain are often characterized by just a few components, then, the resulting TFD $\rho(n, k)$ can be considered as a sparse matrix. Such TFD is in general not full rank and its SVD can be expressed as [51]:

$$\rho(n, k) = \mathbf{U}\mathbf{D}\mathbf{V}^H, \tag{49}$$

where $\mathbf{U}$ is an orthogonal matrix, $\mathbf{V}^H$ is the conjugate transpose of the orthogonal matrix $\mathbf{V}$, and $\mathbf{D} = \text{diag}(\alpha_1, \alpha_2, \cdots, \alpha_{N-r+1}, \cdots, \alpha_N)$ is a diagonal matrix in which the last r singular values are equal to 0 ($[\alpha_{N-r+1}, \alpha_{N-r}, \cdots, \alpha_N] = \mathbf{0}_r^{\mathrm{T}}$). In the presence of white noise, the TFD matrix becomes:

$$\rho(n, k) = \mathbf{U}\mathbf{D}\mathbf{V}^H + \mathbf{N} = [\mathbf{U}_1\mathbf{D}_1\mathbf{V}_1]^H. \tag{50}$$

The TFD matrix is now full rank with its last r singular values being non-zero.

Based on this property, a simple and fast de-noising algorithm can be designed as follows [51]:

- Divide the TFD matrix into sub-blocks $\mathbf{B}_s$.
- Perform SVD on each sub-block $\mathbf{B}_s = \mathbf{U}_s\mathbf{D}_s\mathbf{V}_s^H$ and set to zero the last singular values that are smaller than a threshold ϵ.
- Replace the submatrix $\mathbf{B}_s$ by $\tilde{\mathbf{B}}_s = \mathbf{U}_s\tilde{\mathbf{D}}_s\mathbf{V}_s^H$ ($\tilde{\mathbf{D}}_s$ is the modified diagonal matrix $\mathbf{D}$).
- Reconstruct the denoised matrix $\rho(n, k)$.

Fig. 17 illustrates the enhancement obtained using the aforementioned (t, f) image de-noising algorithm.

Another SVD based de-noising method reduces noise by applying low-pass filtering to singular vectors [52]. The smoothed singular vectors are then used to synthesize a de-noised TFD. Experimental results demonstrate that this method can suppress noise without deteriorating the basic structure of TFDs [52].

b) **Wavelet based methods:** These image de-noising techniques for improving the SNR of (t, f) images are based on the assumption that most of the (t, f) image energy is concentrated in only few coefficients of the Wavelet Transform (WT). The WT is obtained by decomposing a TFD into several bands using a low-pass filter and a high-pass filter [53]. So, a (t, f) image is de-noised by assigning zero value to coefficients below a certain threshold and then inverting the WT. Most of the wavelet de-noising techniques use additive white noise model for noise. This model may be considered as a rough approximation of the noise in QTFDs.

c) **Morphological processing:** These techniques exploit the fact that signal components appear as ridges/curves occupying large connected regions in a (t, f) plane. On the other hand, noise is randomly distributed in a (t, f) plane and appears as small objects or disconnected regions [53]. These small objects can be eliminated from (t, f) images by using the image opening operation [54]. This operation compares a predetermined template (often called the structuring element) with objects in an image; if the size of an object is smaller than that of the template, it is deleted from the image; otherwise it is retained.

d) **Anisotropic diffusion [55]:** Traditional 2D low-pass filtering schemes result in blurring in all directions regardless of object boundaries. The anisotropic diffusion overcomes this limitation by performing edge-preserving image enhancement such that the smoothing is performed on all the points in the (t, f) plane except the points lying near the objects boundaries (i.e., near the boundaries of support regions of the signal components).

4.1.3. Cross-term suppression

Three separate techniques are reviewed below that use image processing techniques.

a) Some techniques use morphological image reconstruction to suppress interference terms in the WVD [11]. Such techniques

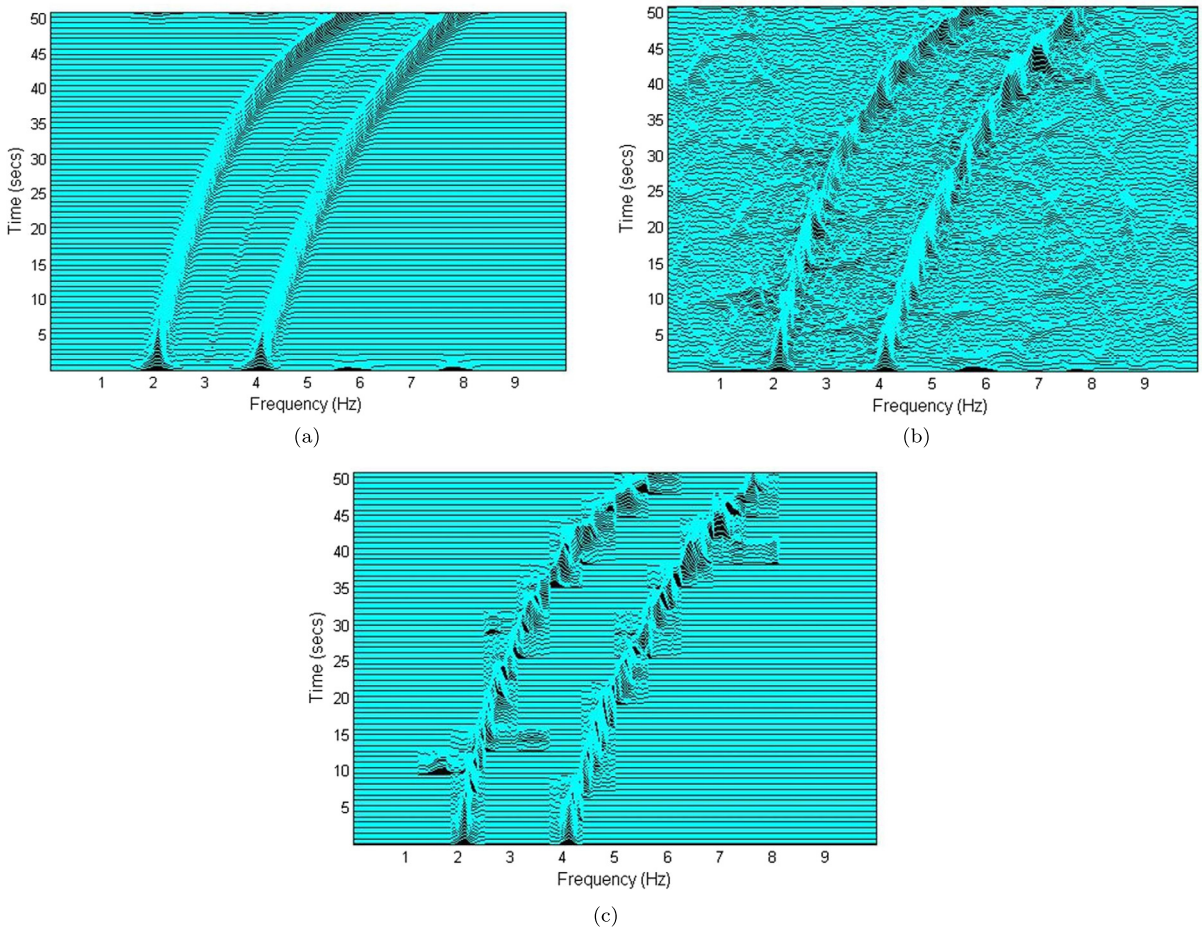


Fig. 17. Illustration of the (t, f) image denoising algorithm. (a) The MBD of a two-component quadratic FM signal. (b) The MBD of the same signal corrupted by additive white Gaussian noise (SNR = 0 dB). (c) The denoised MBD obtained using the algorithm described in [51].

usually employ two images; one is a marker that contains the starting points of the transformation, while the other one is a mask that constrains the transformation. For example, the technique described in [11] processes the WVD, a marker in this case, using the characteristics of the Spectrogram to form a mask. The processing of the WVD starts at peaks. These peaks are then dilated while the thinned Spectrogram constrains the spreading of the peaks. This process continues until the marker stops changing. This technique combines the advantages of the WVD and Spectrogram so that the resulting representation has high energy concentration and is also cross-terms free.

b) Another approach uses a non-linear center affine filter to suppress the cross-terms in the WVD [56]; such a filter is designed to exploit the local statistics of a (t, f) image to adapt its shape. In the regions of high variance it transforms itself into a low-pass filter while in the regions of low variance it adapts itself into an identity operator. Thus it aims to reduce cross-terms while retaining the shape of auto-terms [56].
The limitation of the above cross-term suppression techniques [11,56] is that they fail to give optimum results when cross-terms overlap with auto-terms.

c) A hybrid image and signal processing technique can be used to suppress cross-terms in the WVD even in scenarios when cross-terms overlap with auto-terms [12]. The key stages of this technique are shown in Fig. 18. The technique uses a blurred cross-term-free reduced-interference TFD to locate signal component in a (t, f) plane. The reduced-interference TFD is then segmented using the connectivity criterion discussed in Section 3.3; each segment corresponds to a signal component in the time domain. Fractional Fourier filtering is then applied to perform time-varying filtering so as to extract signal components from the original signal. The extracted components are analyzed separately using the masked WVDs to suppress both inner and outer interference terms. The masked WVDs of the separated signal components are added up to obtain a cross-term free high-resolution TFD. This technique can also be used for estimating the IFs of multi-component signals as the extraction of signal components reduces the problem of IF estimation of multi-component signals to the basic IF estimation of mono-component signals. This technique can be more broadly applied than the first two methods, but it is still limited to signals with non-overlapping (t, f) signal components. The step by step illustration of this third approach for a two-component signal is shown in Fig. 19.

All the above mentioned (t, f) image processing methods are useful tools for the improvement of TFDs readability and energy concentration. However, these methods cannot overcome the fundamental resolution limitation of QTFDs in dealing with closely placed signal components. The following subsection presents a different approach combining directional filtering and image processing techniques for defining a high-resolution TFD that can better deal with such a situation and resolve closely placed signal components.

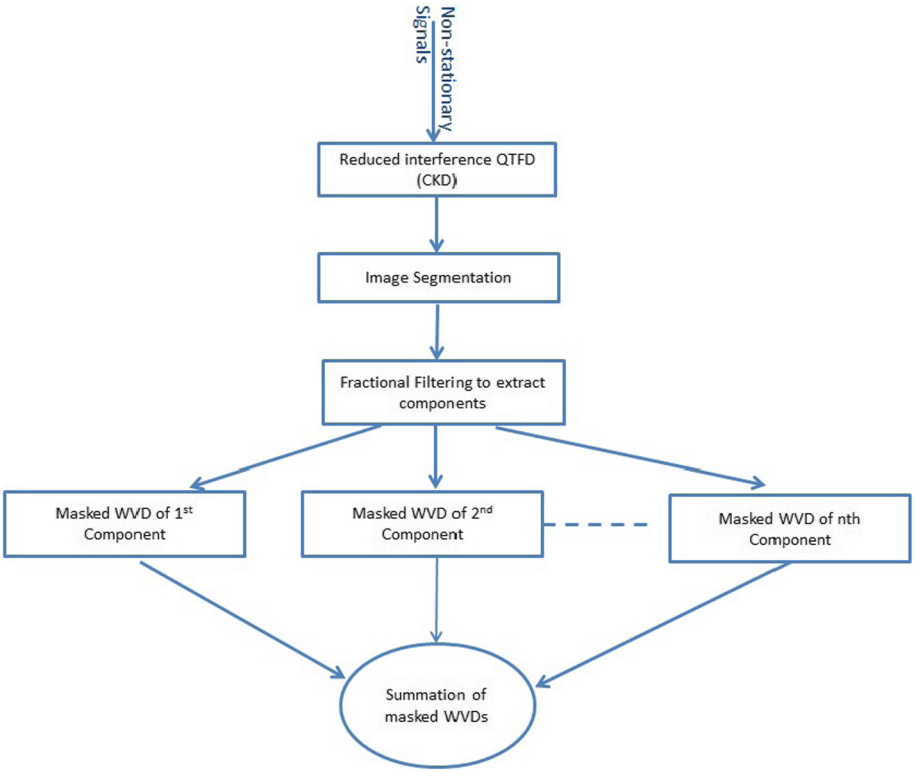


Fig. 18. Flow chart of the TFD resolution enhancement technique [30].

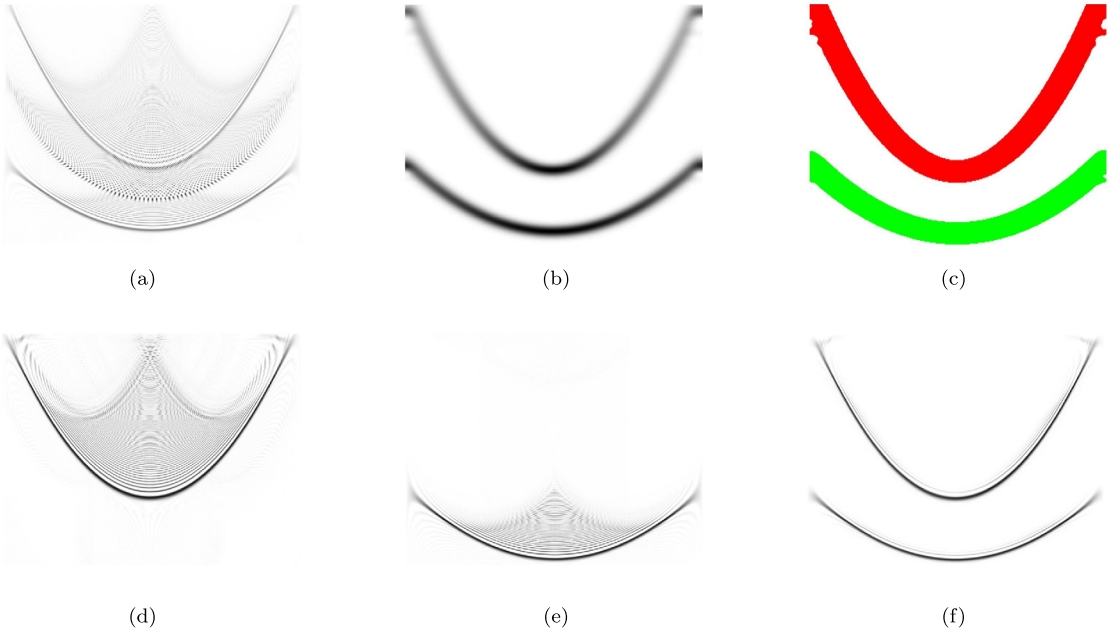


Fig. 19. Step by step illustration of a hybrid approach combining image and signal processing techniques to reduce cross-terms [12]. (a) The WVD. (b) The reduced interference QTFD. (c) (t, f) image segmented such that each segment corresponds to a signal component. (d) The WVD of the first signal component (extracted using fractional filtering). (e) The WVD of the second component (extracted using fractional filtering). (f) The desired (t, f) representation obtained by first masking the WVDs of the extracted signal components with Spectrogram and then adding them, thus illustrating the image processing approach to improve TFDs' resolution.

4.2. Time–frequency image enhancement using adaptive directional filtering

The performance that can be achieved by combining a (t, f) approach with an image processing approach can be further enhanced by using an adaptive directional filtering based (t, f) image enhancement technique. This refinement includes adapting the direction of a (t, f) kernel at each (t, f) point on the basis of the direction of energy concentration resulting, in essence, in the definition of another high-resolution TFD.

4.2.1. Review of signal dependent TFDs with automatic parameter estimation

Most TFDs require optimization of their parameters to obtain a clear and accurate (t, f) representation with the best possible resolution; e.g. incorrect selection of window length for the Spectro-

gram can result in a significantly different visual representation to the correct representation (see [33], p. 39). So, there is a need for some kind of automatic procedure for the selection of parameters in applications requiring real-time computations. Most of the existing adaptive TFDs optimize the parameters of their kernels, either globally or locally, to maximize a certain measure representing the quality of a TFD [57–60]. Some of the commonly used quantitative measures for assessing TFD quality include the Kurtosis (Ratio of norms) [61], the Renyi Entropy [60], and the energy concentration measure [57]. These and other measures are described in detail in Section 5.3. For any given quality measure, the parameters of TFDs can be optimized by either using an exhaustive search technique or employing more computationally efficient methods such as the steepest decent recursive algorithm [57].

4.2.2. Design of the adaptive TFD

Previous studies have reported that for underspread signals, directional filtering along the major axis of the auto-terms can significantly reduce the cross-terms without affecting the resolution of the auto-terms [32,62]. These schemes naturally require prior information about the angle formed by the auto-terms and the time axis. The angle formed by the auto-terms and the time axis can be estimated from the moments of the fractional Fourier transform [63] for signals whose auto-terms have only one direction of energy concentration in a (t, f) plane. In general, there are several possible directions of energy concentration in the (t, f) domain and this would require separate filters for all these directions at a local level. Another solution is to design a post-processing technique to further improve high-resolution TFDs by adapting the direction of a smoothing kernel locally for each point in the (t, f) plane. Such adaptive kernel TFD can be expressed as an extension of Eq. (6), i.e.:

$$\rho_{\text{adapt}}(t, f) = \gamma_{\theta(t,f)}(t, f) \underset{(t,f)}{**} \rho(t, f), \tag{51}$$

where $\rho_{\text{adapt}}(t, f)$ is the adaptive directional TFD (ADTFD), $\gamma_{\theta(t,f)}(t, f)$ is an adaptive kernel whose direction is adapted at each point in the (t, f) plane, and $\rho(t, f)$ is the QTFD used to start the process. A directional smoothing kernel should have the following properties:

- The kernel should have the maximum response when aligned with ridges. This implies that the kernel should have low-pass characteristics along its major axis.
- The kernel output should be zero for non-ridge points as otherwise spectral leakages would be observed at (t, f) points where no signal is present.

The second-order partial derivative of a Directional Gaussian Filter (DGF) is selected as a directional kernel as it fulfills the above properties. This filter has a maximum response when parallel to ridges because of low-pass characteristics along its major axis and its response reduces to zero as the filter is orthogonal to ridges because of differentiation along its minor axis [64–66]. It is defined as:

$$\gamma_{\theta(t,f)}(t, f) = \frac{\partial^2[e^{-((at_\theta)^2+(bf_\theta)^2)}]}{\partial f_\theta^2},$$
$$t_\theta = t\cos(\theta) + f\sin(\theta), \quad f_\theta = f\cos(\theta) - t\sin(\theta), \tag{52}$$

where θ is the rotation angle with respect to the time-axis, a and b control the spread of the DGF along the time and frequency axis respectively. The design criterion to choose the direction of the adaptive kernel for each point in the (t, f) plane is based on the following observations [67]:

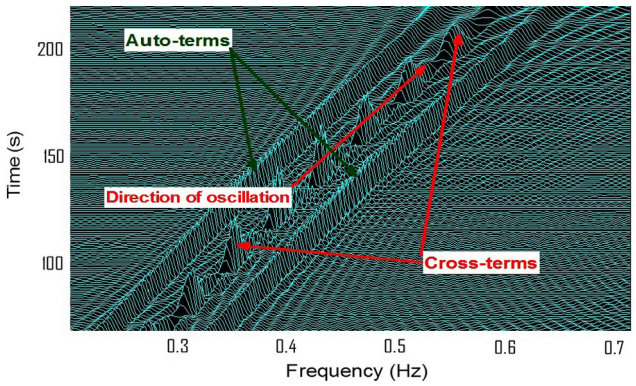


Fig. 20. Illustration of the direction of the auto-terms as well as the direction of the oscillation of cross-terms in the WVD of a signal composed of two LFM components.

1. Cross-terms are oscillatory in the direction of their major axis, but auto-terms are not, as illustrated in Fig. 20.
2. Filtering along the major axis of the auto-terms does not blur a TFD.
3. Both auto-terms and cross-terms appear as ridges in a QTFD.

The implication is that smoothing (low-pass filtering) along the major axis of cross-terms will suppress these cross-terms as they are oscillatory along their major axis (and they are localized away from the origin in the ambiguity domain) while smoothing along the major axis of the auto-terms will not affect their resolution [67]. The direction of the smoothing filter for each (t, f) point is adapted to the direction of the major axis of ridges of a magnitude squared QTFD. The reason is that cross-terms and auto-terms of a QTFD appear as ridges when the squared magnitudes are used because the squared magnitudes of cross-terms have non-oscillatory characteristics in the (t, f) domain due to the squaring operation. The ridge direction (i.e., the direction of maximum energy of a TFD) is estimated by maximizing the correlation of the DGF with the magnitude squared QTFD; expressed as:

$$\hat{\theta}_{(t,f)} = \arg\max_{\theta}\left[\left|\int_{\mathbb{R}}\int_{\mathbb{R}} |\rho(t-t', f-f')|^2 \gamma_{\theta(t,f)}(t', f')\mathrm{d}t'\mathrm{d}f'\right|^2\right] \tag{53}$$

In the above equation, $\rho(t, f)$ is squared to remove the oscillation in the cross-terms as discussed above; on the other hand, $\gamma_{\theta(t,f)}(t, f)$ is not squared as, otherwise, it would change the shape of the DGF filter. Eq. (53) can be numerically solved by performing the following steps:

- Convolve a magnitude squared TFD with K directional filters such that the angle of each filter is given by:

$$\theta_k = \frac{2\pi k}{K}, \quad k = 1, \cdots, K. \tag{54}$$

- For each (t, f) point, choose the angle that maximizes the magnitude of a directionally smoothed TFD.

$$\theta_{t,f} = \frac{2\pi}{K}\arg\max_{k}\left|\gamma_{\theta_k}(t, f) \underset{(t,f)}{**} \left|\rho(t, f)\right|^2\right|^2. \tag{55}$$

The block diagram for the implementation of the proposed ADTFD is shown in Fig. 21.

4.2.3. Performance evaluation

In order to evaluate the performance of the ADTFD, let us consider three multi-component signals with the characteristics that their TFD auto-terms are parallel neither to the time axis nor to

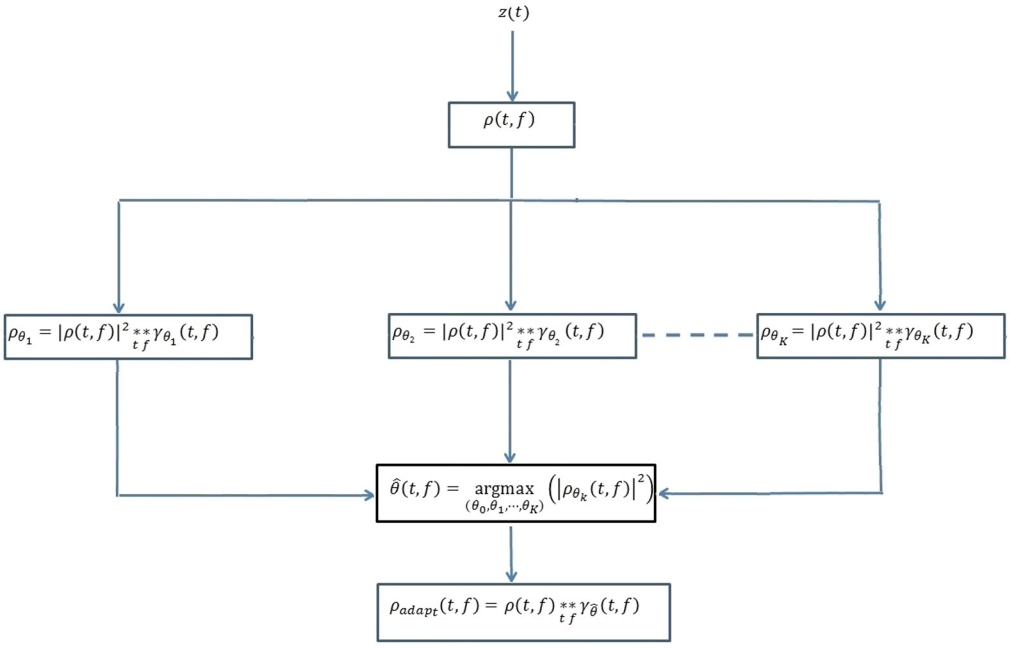


Fig. 21. Implementation of the adaptive kernel TFD. (Note that $\theta_k = 2\pi k/K$, where $k = 1, 2, \ldots K$.)

Table 2
Resolution performance of selected TFDs for a multi-component signal composed of two parallel LFM signals.

TFD	Optimal kernel parameters	P
CKD	$c = 1$, $E = 0.12$, $D = 0.05$	0.8919
Adaptive TFD	$a = 2$, $b = 20$	0.9648
S-Method	Hamming window length = 4 s	0.8778
Extended Modified B distribution	$\alpha = 0.18$, $\beta = 0.06$	0.8376
Modified B-distribution	$\beta = 0.16$	0.7628

the frequency axis. Most separable kernel TFDs fail to achieve a high (t, f) resolution for these signals because of the absence of the (t, f) energy concentration direction parameters in their design. Such parameters relate to the presence and localization of the auto-terms.

a) **Two parallel LFM signals:** The multi-component signal composed of two parallel LFM, described in Section 2.3.4, is used, for consistency and continuity, to evaluate the performance of the DGF adaptive (t, f) image enhancement technique, as compared with the CKD and S-method, as they are found in this study to be the best performing QTFDs in terms of resolution and cross-term suppression capabilities. Fig. 22 illustrates the results. The performance of the proposed TFD is evaluated using the Boashash–Sucic's criterion [38], with resulting P values recorded in Table 2. The proposed adaptive TFD outperforms the CKD by 7% in terms of its ability to retain energy concentration of auto-terms while suppressing cross-terms.

b) **Multi-component signal composed of signal components of varying amplitudes and frequency rates:** let us consider a multi-component signal composed of two LFMs, one Gaussian atom and one tone, expressed as:

$$s(t) = \begin{cases} s_1(t) + s_2(t) + s_3(t) + s_4(t), & 0 \le t \le T, \\ 0, & \text{otherwise}, \end{cases} \tag{56}$$

where the signal components have the characteristics shown below.

Signals	$s_1(t)$	$s_2(t)$
Formula	$\cos(2\pi(0.0006t^2 + 0.16t))$	$\cos(2\pi(0.0006t^2 + 0.1t))$
Signals	$s_3(t)$	$s_4(t)$
Formula	$0.75\cos(0.08\pi t)$	$2e^{-0.0025(t-175)^2}\sin(0.25\pi t)$

The IFs of the signal components are,

IF	$f_1(t)$	$f_2(t)$	$f_3(t)$	$f_4(t)$
Formula	$0.0012t + 0.16$	$0.0012t + 0.1$	0.04	0.125

The sampling frequency is 1 Hz and the signal duration is $T = 255$ s for the signal components $s_1(t), s_2(t)$ and $s_3(t)$. The corresponding effective time duration T_e and effective bandwidth B_e of $s_4(t)$ are given by $T_e = 20\sqrt{\pi/2}$ s and $B_e = 0.05/\sqrt{2\pi}$ Hz (see [68] for these definitions). The total signal has more than one direction of energy concentration and the distribution of the signal energy along each direction is not uniform. Given such complexity, this signal is difficult to analyze using existing fixed-kernel TFDs because these methods do not adapt the kernel locally at each (t, f) point and any particular choice would be adapted only to one component, but would blur the others. The performance of the DGF (t, f) image enhancement technique is compared with the CKD and another adaptive TFD defined earlier [58,59] with results as shown in Fig. 23. The ADTFD gives high energy concentration for all signal components, whereas the CKD and the adaptive TFD [59] fail to concentrate energy for the weak signal component i.e., the tone as observed in the left-hand side of Fig. 23.

c) **Real-life EEG signal:** The previous experiment is repeated using a real-life EEG signal described in Section 2.3.4. Fig. 24 shows that the ADTFD gives high energy concentration for the low-amplitude signal components appearing e.g., above 5 Hz, whereas other TFDs fail to concentrate energy for the weak signal components.

The advantage of QTFDs such as the CKD is that they are useful for most non-stationary signals with different performance depending on the type of signal analyzed. In scenarios when such methods fail to give result with enough resolution, then image processing

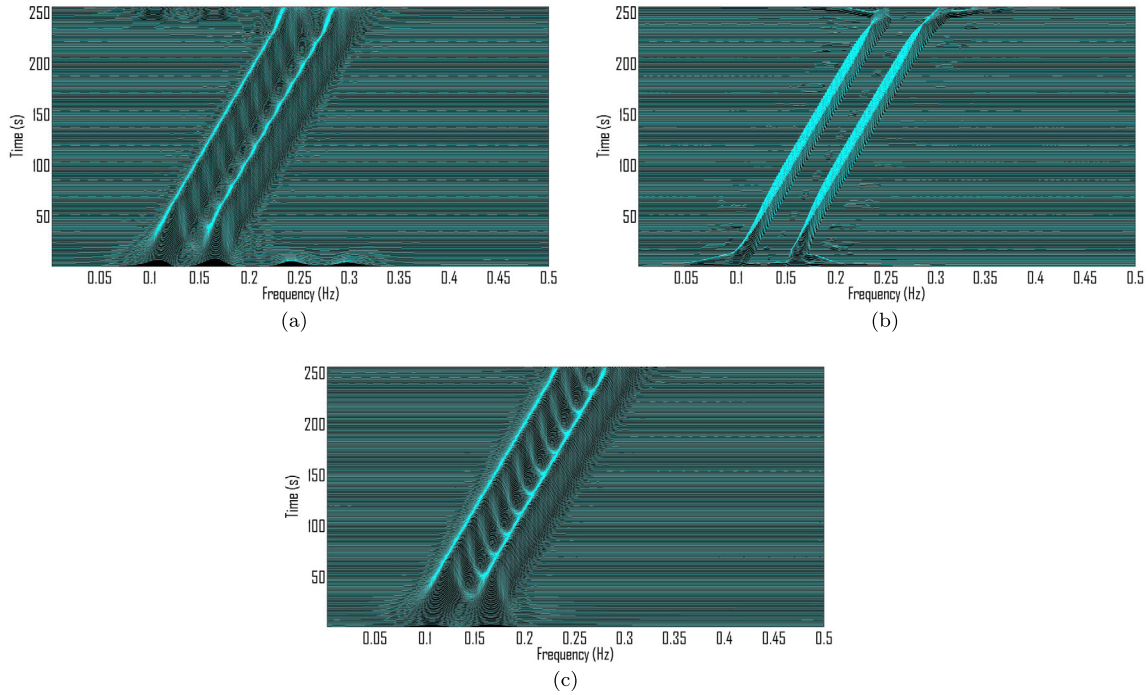


Fig. 22. (t, f) representations of two parallel LFM signals. (a) The CKD ($c = 1$, $E = 0.12$, $D = 0.05$). (b) The adaptive (t, f) image enhancement by DGF ($a = 2$, $b = 20$) and (c) the S-Method (Hamming window length = 4 s).

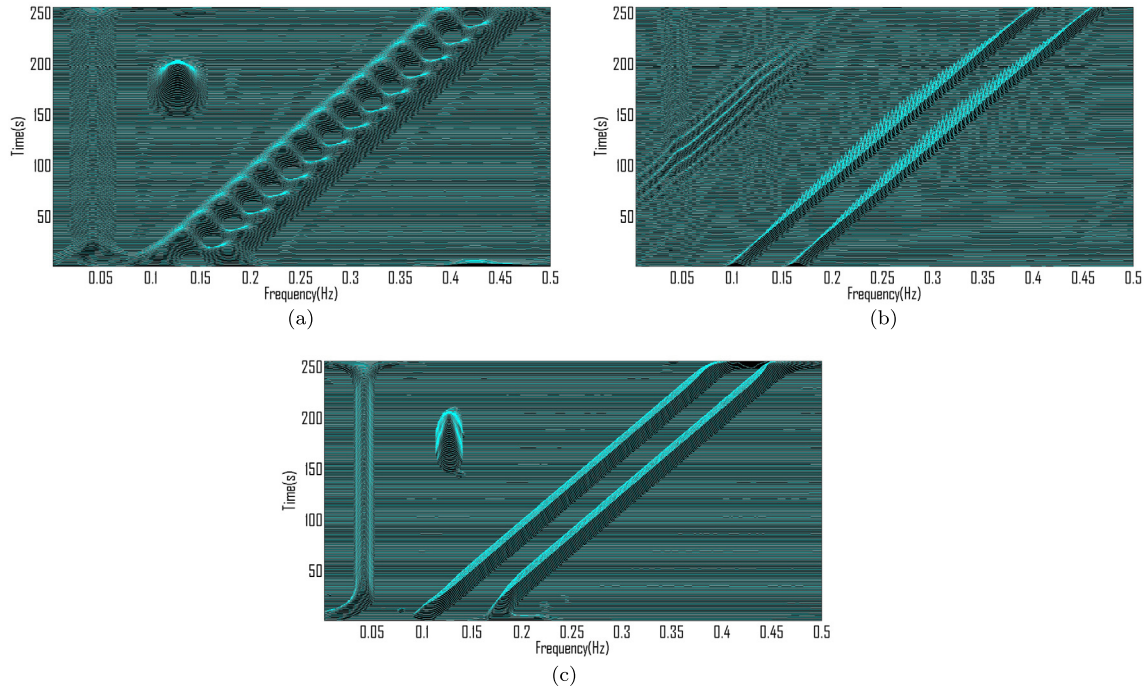


Fig. 23. The (t, f) representations of a multi-component signal composed of signal components of varying amplitudes and chirp rates. (a) The CKD ($C = 0.1$, $E = 0.065$, $D = 0.065$). (b) The Adaptive kernel TFD [59]. (c) The ADTFD ($a = 3$, $b = 8$).

based adaptive techniques can be used as shown in Figs. 22, 23 and 24. These examples illustrate the possible advantage to use image processing techniques such as directional filtering for defining high-resolution cross-term free TFDs. The additional load for doing so is assessed in the next section.

4.2.4. Computational cost of the adaptive TFD

The computation of the adaptive TFD involves the following steps.

1. Computation of a QTFD: The computational cost of this step is $O(NM \log M)$ as it requires N FT operations for a signal of length N and the computational cost of each FT is $O(M \log M)$, where M is the total number of frequency domain samples [69].
2. Computation of an adaptive kernel: The computation of an adaptive directional kernel requires the estimation of an optimal rotation angle for each point in a (t, f) plane. It can be implemented by convolving a QTFD with K directional kernels

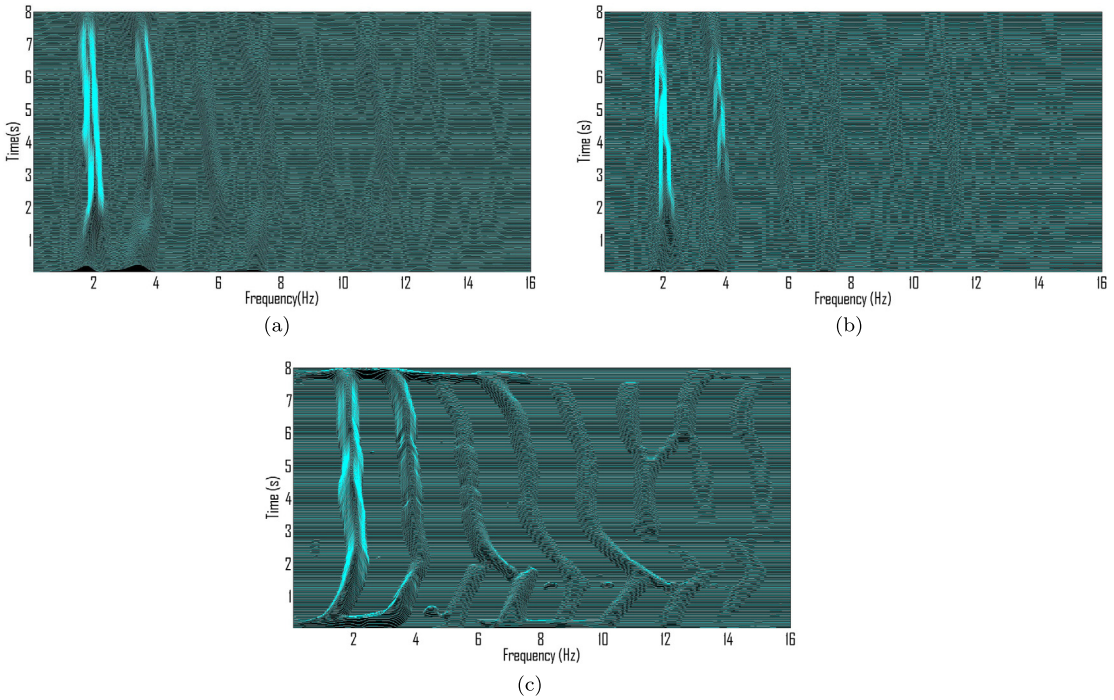


Fig. 24. The (t, f) representations of a real-life EEG signal. (a) The CKD ($c = 1$, $D = 0.2$, $E = 0.05$). (b) The Adaptive kernel TFD [59]. (c) The ADTFD ($a = 3$, $b = 8$).

and choosing the one with maximum response at each (t, f) point. The computational cost of this step is $O(KNM)$ because of the convolution of a QTFD with K directional kernels; given that the computational cost of single convolution operation is $O(NM)$ as the required number of multiplications and additions is proportional to the total number of points in a (t, f) plane.

3. Convolution of a QTFD with an adaptive kernel: The computation cost of this step is $O(NM)$ because it involves only a single convolution.

The total computational cost of this algorithm is equal to the sum of the computational costs of all the above mentioned steps, i.e., $O(NM \log M + KNM + NM)$ or $O(NM \log M + (K + 1)NM)$. The computational load can be significantly reduced by exploiting the fact that (t, f) matrices are sparse so most of the (t, f) points do not need filtering at all.

This section has illustrated that the image processing based methods can be used as post-processing operations to enhance QTFDs. Such post-processing operations not only improves the readability of QTFDs for signal analysis, but also helps in extracting highly discriminating features for pattern recognition applications. A methodology for defining such (t, f) features from high-resolution TFDs is discussed in the next section.

5. Time–frequency feature extraction

Signal feature extraction is a key stage of any overall scheme for pattern recognition and classification of abnormalities or, more generally for any machine learning design and algorithm that requires automatic decision making. The spectral characteristics of a signal are traditionally used to obtain specific details about the underlying signal. Such frequency-domain features [26] include spectral flatness, spectral flux, entropy and average frequency. In the case of non-stationary signals with spectral content varying with time, such frequency-domain features may fail to give sufficient relevant discriminating information. The following subsections discuss in detail the (t, f) extension of some selected frequency-domain features, as an illustration of the (t, f) feature extraction methodology.

5.1. Time–frequency flatness

The (t, f) flatness feature extends the notion of spectral flatness; it is used to measure the relative concentration of energy distribution in a TFD or conversely, its spread. A high (t, f) flatness indicates that the energy is uniformly spread in the TFD while a low (t, f) flatness indicates that the energy is concentrated in specific regions within the TFD. This feature helps, e.g., to discriminate TFDs of signal components embedded in noise from TFDs of pure noise.

In order to derive the mathematical formulation for the (t, f) flatness, let us first consider the definition of the spectral flatness. The spectral flatness is the geometric mean of the FT of a signal normalized by its arithmetic mean; expressed as [70]:

$$\mathrm{SF} = M \frac{(\Pi_{k=1}^{M} |Z[k]|)^{1/M}}{(\sum_{k=1}^{M} |Z[k]|)}. \tag{57}$$

The (t, f) extension of the spectral flatness can then be defined by replacing the spectrum of the signal by its TFD. Eq. (57) then becomes the geometric mean of a TFD normalized by its arithmetic mean, which can be expressed as:

$$\mathrm{TF}_{\mathrm{Flatness}} = \frac{(\Pi_{n=1}^{N} \Pi_{k=1}^{M} \rho[n, k])^{\frac{1}{NM}}}{\frac{1}{NM} \sum_{n=1}^{N} \sum_{k=1}^{M} \rho[n, k]}. \tag{58}$$

The (t, f) flatness assumes a zero value even if there is a single zero in a TFD. So, in practical implementations all zeroes of a TFD are replaced by very small values (e.g., epsilon in Matlab).

5.2. Time–frequency flux

The (t, f) flux extends the stationary spectral flux measure to non-stationary signals. The spectral flux measures the rate of change of the frequency content of a signal with time, whereas the (t, f) flux measures the rate of change of the signal energy in

the (t, f) plane both along time and frequency axes. It is expressed as:

$$\mathrm{TF}_{\mathrm{FLUX}} = \sum_{n=1}^{N-l} \sum_{k=1}^{M-q} \left| \rho[n+l, k+q] - \rho[n,k] \right|, \tag{59}$$

where l and q are predetermined values that depend on the rate of change of signal energy in a (t, f) plane. The values of the parameters used in this study are: $l = 1$ and $q = 1$. This measure can be used to detect seizures in EEG signals as previous studies indicate that the energy of the newborn EEG seizure signals varies slowly along both time and frequency axes. This is inferred from the fact that these signals can either be modeled as piecewise LFM signals with slowly changing frequency or as train of impulses [1]. On the other hand, the EEG background appears as a random pattern. It follows that the (t, f) flux is expected to have a lower value for EEG seizure signals and a larger value for EEG background.

5.3. Time–frequency measures of complexity

A measure of complexity can be defined by interpreting a TFD as a quasi-probability distribution function, and using entropy measures as characteristics. The interpretation is that a highly concentrated TFD with a small number of components has a lower entropy than a signal with a large number of signal components. Some of the entropy measures used in previous studies are defined below and compared in terms of their relevance to the main objectives of this study.

1. Time-Frequency Shannon Entropy (TFSE): It is defined as ([33], p. 299):

$$\mathrm{TFSE} = -\sum_{n=1}^{N} \sum_{k=1}^{M} \rho[n,k] \log_2\big(\rho[n,k]\big). \tag{60}$$

The major limitation of the (t, f) Shannon entropy defined above is that it cannot be used for the majority of QTFDs as these can take negative values, since the above expression is then not valid (log of the TFD). To remedy this, the Renyi Entropy was used instead ([33], p. 298).

2. Time-Frequency Renyi Entropy: It is defined as ([33], p. 298):

$$\mathrm{TFRE} = \frac{1}{1-\alpha} \log_2 \left(\sum_{n=1}^{N} \sum_{k=1}^{M} \rho^{\alpha}[n,k] \right) \tag{61}$$

with $\alpha > 2$; for $\alpha = 2$, the above measure becomes equal to zero for a normalized TFD (i.e., $\sum_{n=1}^{N} \sum_{k=1}^{M} \rho^{\alpha}[n,k] = 1$ for $\alpha = 2$). The (t, f) Renyi Entropy overcomes the limitations of the TFSE by applying the log operator to $\sum_{n=1}^{N} \sum_{k=1}^{M} \rho^{\alpha}[n,k]$, where $\sum_{n=1}^{N} \sum_{k=1}^{M} \rho^{\alpha}[n,k] \geq 0$ for any valid TFD. The (t, f) Renyi entropy is a useful feature for the classification of non-stationary signals because of its sensitivity to the number of signal components, their time duration and bandwidth, and their amplitude ratios [71].

3. Time–frequency normalized Renyi entropy (TFNRE): Zero-mean cross-terms are reduced in the (t, f) Renyi entropy for odd values of α because of the summation operation, so the (t, f) Renyi entropy cannot discriminate a high-resolution TFD with significantly reduced cross-terms from a high-resolution TFD without any suppression of cross-terms. This limitation of the (t, f) Renyi entropy can be overcome by first normalizing the TFD with the sum of its absolute values so that cross-terms have an overall effect of reducing the TFNRE. The TFNRE is thus preferred to evaluate the resolution performance of a TFD as it can differentiate between two TFDs of the same resolution but different levels of cross-term suppression. It is defined as ([33], p. 299, [72]):

$$\mathrm{TFRNE} = -\frac{1}{2} \log_2 \left(\sum_{n=1}^{N} \sum_{k=1}^{M} \left[\frac{\rho[n,k]}{\sum_{n=1}^{N} \sum_{k=1}^{M} |\rho[n,k]|} \right]^{\alpha} \right). \tag{62}$$

Some other complexity measures that are not related to entropy, but can be used to discriminate a low resolution TFD from a high-resolution TFD, are mentioned below.

- Kurtosis (or Ratio of norms): It is defined as the ratio of the fourth-order moment of a TFD with the squared second order moment [61]:

$$\mathrm{RN} = \frac{\sum_{n=1}^{N} \sum_{k=1}^{M} \rho^4[n,k]}{\left(\sum_{n=1}^{N} \sum_{k=1}^{M} \rho^2[n,k]\right)^2}. \tag{63}$$

  This kurtosis is used to estimate the "peakedness" of TFDs by calculating the above ratio. A high value of the ratio of norms indicates that the TFD is highly concentrated while a low value indicates that the TFD is poorly concentrated. In many real-life situations, signal components can have significant variations in their relative amplitudes. For such signals, a TFD optimized based on the kurtosis can achieve a high energy concentration for strong components at the expense of poor energy concentration for weak signal components.
- Energy concentration measure: One approach to estimate the spread of the energy distribution in the (t, f) domain for any given TFD is to measure its support region (e.g., the number of non-zero values). A highly concentrated TFD will have a smaller region of support as compared to a blurred TFD. The following energy concentration measure can be used for estimating this support region for a normalized TFD (i.e., $\sum_{n=1}^{N} \sum_{k=1}^{M} \rho[n,k] = 1$) [57].

$$\mathrm{EC} = \left| \sum_{n=1}^{N} \sum_{k=1}^{M} \left| \rho[n,k] \right|^{\frac{1}{q}} \right|^{q}, \quad q > 1 \tag{64}$$

  Unlike the kurtosis, TFDs optimized on the basis of this measure do not suffer from the problem of poor energy concentration for weak signal components [57].

In the reminder of this study, the normalized Renyi entropy is used as the results confirm that it gives better classification results as compared to the Renyi entropy. There are other measures of complexity that have been defined and used, e.g., in [73]; but they are not considered here due to space, scope and relevance limitations. Note that this section has illustrated a methodology to define new (t, f) features by extending frequency-domain features to the (t, f) domain. A similar methodology can also be used to define new features by extending time-domain features to the (t, f) domain [26]. A list of such features is given in Appendix D.

6. Results, discussion and further insights

Receiver Operating Characteristics (ROC) analysis is used to compare the performance of the (t, f) features with the corresponding frequency-domain features for the detection of seizure activity [74]. The area under ROC curve (AUC) is used as a performance metric. The AUC of a feature is the probability that the feature would have a higher value for a randomly chosen positive example than for a randomly chosen negative example [74]. A set of 200 epochs of newborn EEG background and a set of 200 epochs of newborn EEG seizure are randomly extracted from a database for this study. The length of each segment of database is 256 samples (with 32 Hz sampling frequency), corresponding to a duration

Table 3
AUC analysis of the frequency-domain features.

Features	AUC values
Spectral Flatness	0.6632
Spectral Entropy	0.5901
Spectral Flux	0.5483

Table 4
AUC analysis of the (t, f) features obtained by extending the frequency-domain features to the (t, f) domain. These features are extracted from the WVD, Spectrogram, ADTFD, EMBD, and CKD.

Features	ADTFD	CKD	EMBD	WVD	Spectrogram
(t, f) Flatness	0.9092	0.7997	0.7610	0.6922	0.6254
(t, f) Entropy	0.6631	0.7207	0.7228	0.6157	0.7200
(t, f) Flux	0.7792	0.7103	0.5490	0.7344	0.6280

of 8 s. The relevant database is described in [75], and the relevant components used in this study are provided as Supplementary material.

In order to extract EEG features, each segment is analyzed by selected TFDs; these are the Spectrogram (SPEC), WVD, EMBD, CKD, and ADTFD. The parameters that lead to the highest AUC values are selected for each TFD by performing an experimental non-exhaustive search. The selected smoothing parameters for the EMBD are $\alpha = 0.05$ and $\beta = 0.05$. The selected window length L for the SPEC is the Hamming window of $L = 75$ samples. The selected parameters for the CKD are $c = 1$, $D = 0.04$ and $E = 0.04$. The selected parameters for the ADTFD are $a = 3$ and $b = 8$. Note that we have not used the MBD and CSK-TFD for performance evaluation as they are special cases of the EMBD and CKD as discussed in Section 2. The performance of (t, f) features is compared with frequency-domain features using an AUC analysis. Table 3 shows AUC values for the frequency-domain features while AUC values for the corresponding (t, f) extensions are indicated in Table 4.

The key observations regarding the AUC analysis of signal related features as observed from Tables 3 and 4 are given below.

1. (t, f) features, which are extensions of the frequency-domain features, lead to higher AUC values than corresponding frequency-domain features for all TFDs with the exception of the Spectrogram. The Spectrogram has lower AUC value for the (t, f) flatness than the corresponding frequency-domain feature i.e., the spectral flatness.
2. The (t, f) flatness is the best performing (t, f) feature with AUC value of 0.9092 for the adaptive TFD.
3. The adaptive TFD is the most suitable TFD for extracting (t, f) flatness and (t, f) flux because of higher AUC values of these features for the adaptive TFD; the CKD is the second best TFD for (t, f) flatness while the WVD is the second best TFD for (t, f) flux.
4. The EMBD is the best TFD for extracting the (t, f) entropy followed by the CKD.
5. The poor performance of the Spectrogram for extracting the (t, f) flatness can be explained by the sensitivity of the Spectrogram to its window length, which makes it difficult to find an optimal window for a large class of signals.

6.1. Discussion and interpretation of results

The experimental findings indicate that the performance of frequency-domain features can be significantly improved by extending them to the (t, f) domain. It is important to use high-resolution TFDs for extracting (t, f) features as the best performing (t, f) features are extracted from the ADTFD, which is also the best performing TFD in terms of resolution and cross-term suppression capabilities. These observations indicate that, for a given TFD, the performance of a (t, f) abnormality classification technique can be directly related to the resolution of a TFD; hence the need to define data-dependent high-resolution QTFDs suitable for specific classes of signals. The improvements obtained as a result of (t, f) translation of features can be explained as follows:

- The Spectral flatness and Spectral entropy measure the uniformity of energy distribution in the frequency-domain. They can discriminate a narrowband signal from a wideband signal, but they cannot discriminate two wideband signals; e.g. an LFM signal cannot be discriminated from white noise. The (t, f) flatness and (t, f) entropy can discriminate random noise from signals whose energy is concentrated in the (t, f) domain. The energy of EEG seizure signals is usually concentrated along the IFs of their components, whereas the energy of EEG background is usually spread in the (t, f) domain. Therefore, seizure signals have lower (t, f) flatness and (t, f) entropy as compared to background signals.
- The better performance of the (t, f) entropy can be explained by the observation that EEG seizure signals have slow variation of signal energy along both time and frequency axis, whereas the background has a random distribution of signal energy. This implies that for seizure signals there is a little variation of signal energy along the diagonal axis. The (t, f) flux exploits this information by measuring the total variation in the signal energy along the diagonal axis, which results in low values for seizure signals and higher values for the background. On the other hand, the spectral flux only measures the variation in spectral content of EEG signals and cannot measure the rate of change of signal energy along the diagonal axis.

In order to extend and apply the proposed (t, f) pattern recognition approach to make it suitable for a wide range of applications and real-time implementation, several key points need further investigation, as described next.

6.2. Additional time–frequency insights to further improve performance

6.2.1. Designing high-resolution adaptive TFDs

The performance of all QTFDs depends on the choice of certain parameters. For example, the shape and size of an analysis window controls the time and frequency resolution of the Spectrogram. These parameters are data dependent and one set of parameters cannot give optimal results for all classes of data. Moreover, it is difficult to optimize a single global kernel for all the points in the TFD as there can be many components in a signal with different directions of energy concentration. In order to automatically select the optimal parameters that lead to a high energy concentration TFD in which cross-terms and noise are reduced, several adaptive methods were proposed [58,61,59,76]. These methods (e.g., [61, 59]) automatically adapt the parameters of the (t, f) kernel at each time instant, but there is a need in some situations to extend these algorithms so that the kernel parameters are optimized locally at each (t, f) point as done in Section 4.2.

6.2.2. Designing multidirectional fixed-kernel TFDs

Separable-kernel TFDs give good performance for signals whose energy is either concentrated along the time axis or the frequency axis in the (t, f) domain or along a direction with small deviation from one of these two axes. However, these TFDs fail to give optimal energy concentration for signals whose energy is concentrated along a direction significantly away from both time axis and frequency axis in the (t, f) domain. In order to obtain high-resolution (t, f) representation for such signals, a rotation parameter can be included in the formulation of smoothing kernels [63]. For exam-

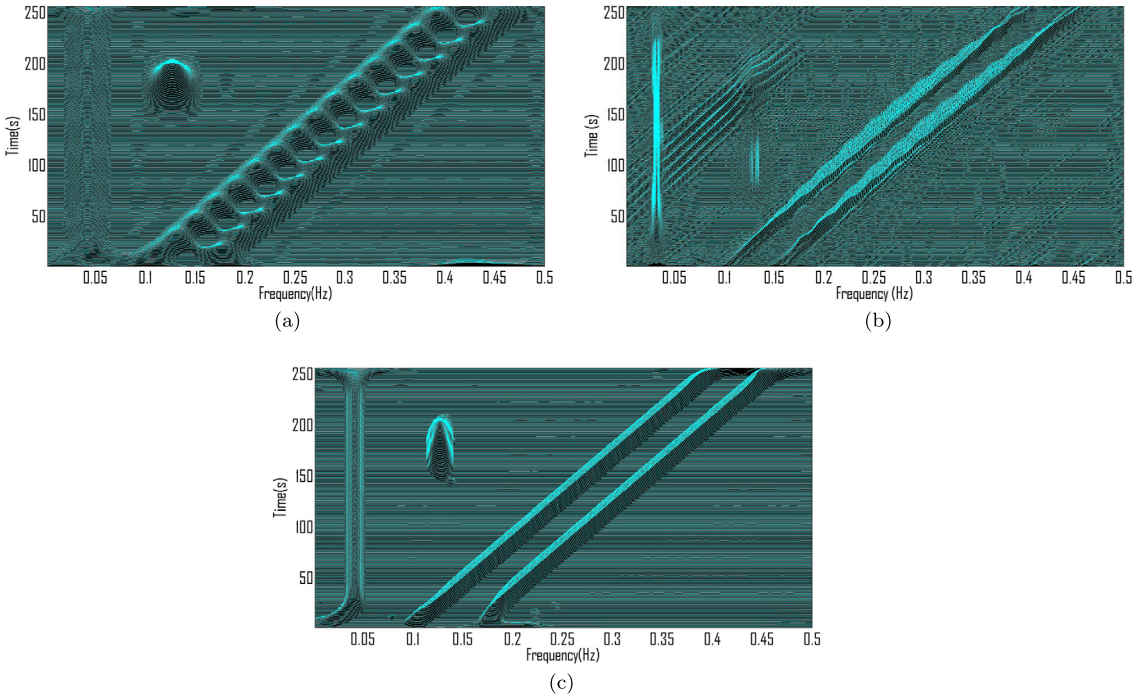


Fig. 25. The (t, f) representations of a multi-component signal composed of signal components of varying amplitudes and chirp rates. (a) The CKD ($c = 0.1$, $E = 0.065$, $D = 0.065$). (b) The multi-directional kernel ($\theta_1 = 0°$, $\theta_2 = 30°$, $c = 0.1$, $D = 0.03$). (c) The ADTFD ($a = 3$, $b = 8$).

ple, an additional parameter θ can be included to allow for the rotation of the CSK, resulting in the following expression.

$$g(\nu, \tau, \theta) = \begin{cases} e^{\frac{c}{(\frac{\nu\cos(\theta)-\tau\sin(\theta)}{D})^2-1}}, & |\cos(\theta)\nu - \sin(\theta)\tau| < D, \\ 0, & \text{otherwise}, \end{cases} \tag{65}$$

where θ is the angle of the kernel with the Doppler axis in the ambiguity domain or the time axis in the (t, f) domain. Rotated (t, f) kernels cannot give ideal representations for signals that have more than one possible directions of energy concentration in the (t, f) domain. Adaptive kernel methods, such as the one defined in Section 4.2.2, are well suited for such signals, but these methods are computationally expensive. One way to analyze signals with multiple directions of energy concentration in the (t, f) domain is to perform smoothing in multiple directions. An example is provided by the following expression which defines a new multidirectional kernel as a summation of a predetermined number of directional kernels (as defined in Eq. (65)):

$$g(\nu, \tau) = \frac{e^c}{N_D} \sum_{i=1}^{N_D} g(\nu, \tau, \theta_i), \tag{66}$$

where N_D is the total number of directional kernels, and θ_i is the angle of the ith kernel.

In order to compare the performance of this multi-directional distribution (MDD) with the other methods, let us use the multi-component signal expressed in Eq. (56). This signal is selected as it has multiple directions of energy concentration in the (t, f) domain, so conventional methods that do not take direction into account cannot be used to optimally analyze this signal. Fig. 25 shows this signal analyzed using the MDD and the two best performing TFDs that are the CKD and adaptive TFD. The MDD gives highest energy concentration for the two parallel LFM signals and tone, but fails to concentrate energy for the Gaussian atom. The lowest region of performance of the MDD kernel concerns the Gaussian atom, which is due to a lack of specific direction. The CKD fails to resolve the close components due to the lack of direction parameter in its formulation. The adaptive TFD, due to the local adaptation of its kernel, has resolved the close components without degrading the energy concentration for the Gaussian atoms. This MDD should therefore be a preferred TFD for non-stationary multi-component and multi-directional signals such as EEG seizures.

6.2.3. Time–frequency image enhancement of TFDs

Some of the state of the art (t, f) image enhancement techniques discussed in Section 4 and others such as [77] can be used as post processing operations. The aim is either to improve the resolution of cross-term-free low resolution TFDs (e.g., the Spectrogram) by using de-blurring methods or to suppress cross-terms in high-resolution TFDs (e.g., the WVD). By including such image enhancement techniques within the overall methodology of abnormality detection, improvements in performance can be obtained in applications such as classification of physiological signals. For example, this study illustrates that the two best (t, f) features (i.e., (t, f) flatness and (t, f) flux) for the detection of seizure activity in EEG signals are obtained from (t, f) images enhanced by the DGF (i.e. the ADTFD). The expense for this result is an additional computational cost, whereas among the non-adaptive methods (i.e., without enhancement) the CKD gives the best performing features and this method is computationally less expensive.

6.2.4. Extraction of geometric features

Signal components appear as pockets of energy concentration in the (t, f) domain or as segments in (t, f) images. In these cases image segmentation techniques can then be used to locate these pockets of energy concentration [78]. Geometric features can then be extracted from these segments to get information such as the number of signal components, their relative positions in the (t, f) plane and the energy distribution of each signal component in the (t, f) plane. This information is critical for abnormality detection using change analysis. For example, in the case of a heart sound signal, the murmur appears as an additional pocket of energy in the (t, f) domain (see Fig. 4 in [5]). So, the number of pockets

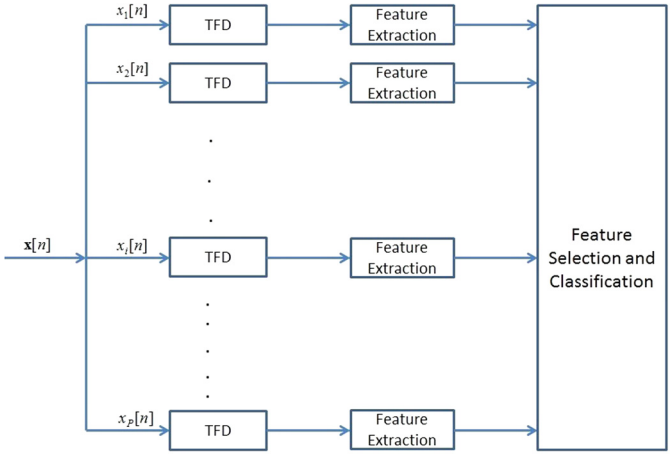


Fig. 26. The proposed block diagram of a multi-channel (t, f) abnormality detection methodology using channel-by-channel approach.

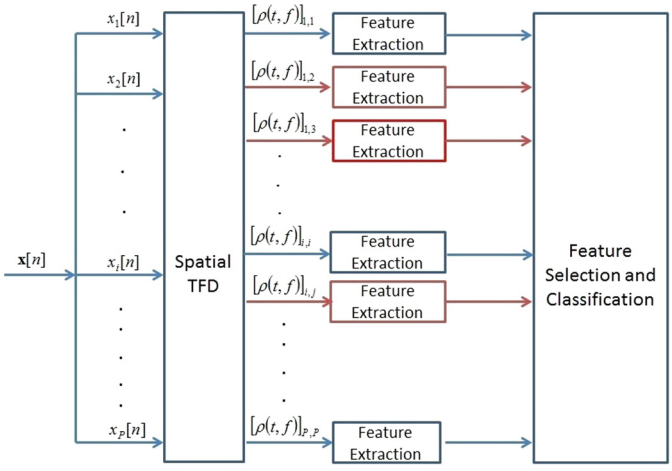


Fig. 27. The proposed block diagram of the multi-channel (t, f) abnormality detection methodology using spatial TFDs. Arrows in red indicate cross-TFDs, while arrows in blue indicate auto-TFDs.

of energy concentration, their location and their geometric properties can be used as features in machine learning algorithms for automatically detecting an abnormality in a heart sound signal.

6.2.5. Extraction of features from multi-channel signals

In many cases, multi-channel and multi-modal recordings are produced, as in the case of physiological signals. Let us consider two approaches for extracting (t, f) features from multi-channel physiological signals, as outlined below.

1. Analyze each channel separately using a high-resolution TFD and then extract features from the TFD as shown in Fig. 26. This approach ignores mutual information among various channels.
2. Analyze a multi-channel signal using a spatial TFD (STFD) and then extract features from the STFD matrix as shown in Fig. 27. For a multi-channel signal $\boldsymbol{x}[\boldsymbol{n}] = [x_1[n], x_2[n], x_3[n], \ldots x_P[n]]$, an STFD matrix for time index n and frequency index k is defined as ([33], pp. 325):

$$\boldsymbol{\rho}[\boldsymbol{n}, \boldsymbol{k}] = \sum_{m=-M}^{M} \sum_{p=-N}^{N} G[n-p, m]\boldsymbol{x}[p+m]\boldsymbol{x}^H \times [p-m]e^{-4j\pi \frac{mk}{M}}, \quad (67)$$

where $G[n, m]$ is a time–lag kernel. Each element of a STFD matrix is defined as:

$$\left[\boldsymbol{\rho}[\boldsymbol{n}, \boldsymbol{k}]\right]_{il} = \sum_{m=-M}^{M} \sum_{p=-N}^{N} G[n-p, m]x_i[p+m]x_l{}^* \times [p-m]e^{-4j\pi \frac{mk}{M}}. \quad (68)$$

The off-diagonal elements of the STFD are the cross-sensor TFDs and the diagonal entries are the standard auto-sensor TFDs of the multi-channel non-stationary signals. Such STFD-based methods can exploit both the non-stationary characteristics of the signals and the spatial diversity provided by the multiple sensors. This is of course at the expense of complexity and a large number of features.

The aforementioned feature extraction approaches can be easily adapted for multi-modal signals by interpreting signals of other modality acquired from other sensors as separate channels.

7. Conclusion

This tutorial review paper describes two key methods and techniques needed to design (t, f) pattern recognition techniques suitable for the classification of non-stationary signals such as those encountered in physiological measurements. The two principal stages include:

1. Design of high-resolution TFDs for better extraction of information defining discriminatory features.
2. Definition of new (t, f) features by extending traditional time-domain or frequency-domain features to (t, f) features such as (t, f) flatness and (t, f) flux.

The performance of the (t, f) features is evaluated using ROC analysis. The findings show that (t, f) features perform significantly better than frequency-domain features; and high-resolution TFDs result in (t, f) features with higher AUC. The other findings of this study with respect to TFD design are:

1. The Spectrogram is easy to use as it depends only on the window size and shape; but it is too sensitive to the window variations. Improvements on the Spectrogram require the use of separable kernel TFDs with more control parameters as in the EMBD, CKD, multi-view TFD, adaptive TFD.
2. LI-TFDs are more suitable for signals that can be modeled as pure tones, i.e., signals with energy concentration parallel to the time axis in the (t, f) domain or small deviation from the time axis; this includes, e.g., some EEG seizure signals.
3. DI-TFDs are more suitable for signals that can be modeled as a train of impulses, i.e., signals whose energy concentration is parallel to the frequency axis in the (t, f) domain; this includes, e.g., spike signals.
4. Separable-kernel TFDs have more degrees of freedom to adapt their smoothing kernels as compared to the Spectrogram, DI-TFDs and LI TFDs. They can be optimized for both signals composed of spikes as well as pure tones.
5. The CKD is the best performing separable-kernel TFD as it has more control parameters to adapt both the shape and size of its smoothing kernel. It is an extension of the EMBD which is itself an extension of the MBD, a modified version of the BD. This best performance is obtained at the cost of greater complexity.
6. The adaptive TFD is an extra refinement that is especially suitable for signals that have more than one direction of energy distribution in the (t, f) plane. It can be applied to the best

high-resolution TFD selected for a large class of signals; its improved performance is obtained at the expense of an additional computational cost.

In order to further improve the performance of existing (t, f) approaches with respect to pattern recognition and machine learning requirements, several additional investigations are needed including:

1. Developing algorithms for the automatic design of data dependent high-resolution TFDs and IF estimation, taking into account directional filtering of TFDs on the basis of energy concentration location criteria.
2. Investigating the further use of directional filtering to extract texture related features.
3. Applying (t, f) image segmentation techniques to extract geometric features to facilitate 1) and 2) [78].
4. Investigating the design of TFDs for compressed sensing [79].

The results from these additional investigations should help bring these techniques to a wider range of applications, and enable advanced machine learning in a (t, f) domain. In particular, the above methodologies and results presented in this paper indicate that in the context of detecting abnormalities in newborn brains, systems that aid experts decision making can be implemented for use in neonatal intensive care units on a trial basis. The same is true for other applications such as condition monitoring and fault detection in a wide range of industrial applications.

The Appendices A–E are provided for completeness and to ensure that the results presented can be reproduced independently by the reader.

Acknowledgments

This research was funded by Qatar Foundation grants NPRP 4-1303-2-517, NPRP 6-885-2-364 and NPRP 6-680-2-282. The authors thank Dr. Brahim Jawad for his valuable technical feedback and Prof. Yimin Zhang for proof checking the draft manuscript. Thanks are also due to the anonymous reviewers for their useful suggestions.

Appendix A. Resolution performance measure used for comparisons of TFDs

The normalized performance measure P introduced in [38] takes into account the proportion of cross-terms in the TFD and provides a measure of resolution of the TFD, as illustrated in Fig. 28. The measure 'P' is defined as:

$$P(t) = 1 - \frac{1}{3}\left\{\left|\frac{A_s(t)}{A_m(t)}\right| + \frac{1}{2}\left|\frac{A_x(t)}{A_m(t)}\right| + \left(1 - \frac{S(t)}{D(t)}\right)\right\} \tag{69}$$

where $A_m(t)$ and $A_s(t)$ are respectively the average amplitudes of the components' mainlobes and sidelobes, $A_x(t)$ is the cross-term amplitude. The parameter $S(t) = (f_2(t) - \frac{B_2(t)}{2}) - (f_1(t) + \frac{B_1(t)}{2})$ represents the component's separation measure and $D(t) = f_2(t) - f_1(t)$ is the distance between the peaks of the two components.

Appendix B. Computational complexity and real-time implementations

The implementation of the abnormality detection methodologies described above involves two stages: offline training and online implementation. The computational complexity of the online implementation is often a major bottle neck in practical realizations of pattern recognition schemes. For the algorithms described above, this involves TFDs computations [69], features extraction, followed by signal classification using a trained classifier. The cost of feature selection is not considered as this step is part of offline training. The computational cost of implementing a TFD is $O(NM\log(M))$, where N is the number of samples of the signal being considered and M is the number of sample points in the FFT computations. The computation of a TFD requires N FT operations (one at each time instant) and each FT is of the order $O(M\log(M))$. The computational complexity of feature extraction in general is $O(NM)$ as there are NM points in the (t, f) plane and the number of mathematical operations required to extract any given feature is proportional to the total number of points. The computational complexity of abnormality detection based on the trained classifier is $O(N_f)$; where N_f is the total number of features. The total computational cost of online computation is therefore $O(NM\log(M) + NM + N_f)$ or simply $O(NM\log(M))$, given that, in the analysis of algorithms, computational costs of higher order is only considered as per general practice [80]. The above analysis indicates that the TFD computation is the major bottleneck in trying to achieve an efficient real-time implementation. Further optimization can be made by using specially designed computational and memory efficient algorithms that have been developed for TFD implementation [69], implementing (t, f) kernels as multi-taper Spectrograms [81], and using efficient hardware implementations such as field-programmable gate arrays.

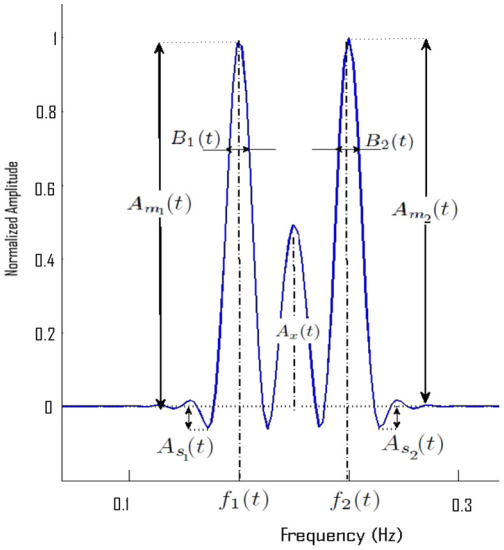


Fig. 28. The two dominant peaks are the signal components; the middle peak is the cross-term, and the other peaks are side lobes.

Appendix C. Computer programs in this study

All simulations have been carried out using Matlab and the TFSAP tool-box. The TFSAP tool-box can be downloaded from the following link. http://espace.library.uq.edu.au/view/uq:211321, http://faculty.qu.edu.qa/boualem/tfsa.aspx, and/or http://www.time-frequency.net. The corresponding code can be found in [82,33].

Appendix D. Time–frequency extension of time-domain features

Table 5 gives the list of (t, f) features that are obtained by translation of time-only features to the (t, f) domain.

Appendix E. Medical data used in this study

Supplementary material related to this article can be found online at http://dx.doi.org/10.1016/j.dsp.2014.12.015.

Table 5
Time–frequency extension of time-domain features.

Feature	Time-domain representation	(t, f) extension
Mean	$m_{(t)} = \frac{1}{N}\sum_{n=1}^{N} s[n]$	$m_{(t,f)} = \frac{1}{NM}\sum_{n=1}^{N}\sum_{k=1}^{M} \rho[n,k]$
Variance	$\sigma^2_{(t)} = \frac{1}{N}\sum_{n=1}^{N} (s[n] - m_{(t)})^2$	$\sigma^2_{(t,f)} = \frac{1}{NM}\sum_{n=1}^{N}\sum_{k=1}^{M} (\rho[n,k] - m_{(t,f)})^2$
Skewness	$\gamma_{(t)} = \frac{1}{N\sigma^3_{(t)}}\sum_{n=1}^{N} (s[n] - m_{(t)})^3$	$\gamma_{(t,f)} = \frac{1}{NM\sigma^3_{(t,f)}}\sum_{n=1}^{N}\sum_{k=1}^{M} (\rho[n,k] - m_{(t,f)})^3$
Kurtosis	$k_{(t)} = \frac{1}{N\sigma^4_{(t)}}\sum_{n=1}^{N} (s[n] - m_{(t)})^4$	$k_{(t,f)} = \frac{1}{NM\sigma^4_{(t,f)}}\sum_{n=1}^{N}\sum_{k=1}^{M} (\rho[n,k] - m_{(t,f)})^4$
Coefficient of variation	$c_{(t)} = \frac{\sigma_{(t)}}{m_{(t)}}$	$c_{(t,f)} = \frac{\sigma_{(t,f)}}{m_{(t,f)}}$

References

[1] B. Boashash, G. Azemi, J.M. O'Toole, Time–frequency processing of nonstationary signals: advanced TFD design to aid diagnosis with highlights from medical applications, IEEE Signal Process. Mag. 30 (6) (2013) 108–119, http://dx.doi.org/10.1109/MSP.2013.2265914.

[2] V. Chen, H. Ling, Joint time–frequency analysis for radar signal and image processing, IEEE Signal Process. Mag. 16 (2) (1999) 81–93, http://dx.doi.org/10.1109/79.752053.

[3] G. Gaunaurd, H. Strifors, Signal analysis by means of time–frequency (Wigner-type) distributions – applications to sonar and radar echoes, Proc. IEEE 84 (9) (1996) 1231–1248, http://dx.doi.org/10.1109/5.535243.

[4] B. Boashash, B. Escudie, Wigner–Ville analysis of asymptotic signals and applications, Signal Process. 8 (3) (1985) 315–327, http://dx.doi.org/10.1016/0165-1684(85)90109-4.

[5] L.D. Avendano-Valencia, J.I. Gondino-Llorente, M. Blanco-Velasco, G. Castellanos-Dominguez, Feature extraction from parametric time frequency representations for heart murmur detection, Ann. Biomed. Eng. 38 (8) (2010) 2716–2732, http://dx.doi.org/10.1007/s10439-010-0077-4.

[6] A.T. Tzallas, M.G. Tsipouras, D.I. Fotiadis, Automatic seizure detection based on time–frequency analysis and artificial neural networks, Comput. Intell. Neurosci. 1 (2007) 1–13, http://dx.doi.org/10.1155/2007/80510.

[7] A.T. Tzallas, M.G. Tsipouras, D.I. Fotiadis, Epileptic seizure detection in EEGs using time–frequency analysis, IEEE Trans. Inf. Technol. Biomed. 13 (5) (2009) 703–710, http://dx.doi.org/10.1109/TITB.2009.2017939.

[8] L.J. Stanković, M. Daković, T. Thayaparan, Time-Frequency Signal Analysis with Applications, Artech House, Boston, 2014.

[9] H. Lee, Z.Z. Bien, A variable bandwidth filter for estimation of instantaneous frequency and reconstruction of signals with time-varying spectral content, IEEE Trans. Signal Process. 59 (5) (2011) 2052–2071, http://dx.doi.org/10.1109/TSP.2011.2113345.

[10] M. Abed, A. Belouchrani, M. Cheriet, B. Boashash, Time–frequency distributions based on compact support kernels: properties and performance evaluation, IEEE Trans. Signal Process. 60 (6) (2012) 2814–2827, http://dx.doi.org/10.1109/TSP.2012.2190065.

[11] S. Gomez, V. Naranjo, R. Miralles, Removing interference components in time frequency representations using morphological operators, J. Vis. Commun. Image Represent. 22 (5) (2011) 401–410, http://dx.doi.org/10.1016/j.jvcir.2011.03.007.

[12] N.A. Khan, I.A. Taj, M.N. Jaffri, S. Ijaz, Cross-term elimination in Wigner distribution based on 2D signal processing techniques, Signal Process. 91 (3) (2011) 590–599, http://dx.doi.org/10.1016/j.sigpro.2010.06.004.

[13] E. Sejdic, I. Djurovic, J. Jiang, Time–frequency feature representation using energy concentration: an overview of recent advances, Digit. Signal Process. 19 (1) (2009) 153–183, http://dx.doi.org/10.1016/j.dsp.2007.12.004.

[14] M. Mboup, T. Adali, A generalization of the Fourier transform and its application to spectral analysis of chirp-like signals, Appl. Comput. Harmon. Anal. 32 (2) (2012) 305–312, http://dx.doi.org/10.1016/j.acha.2011.11.002.

[15] B. Boashash, Time–frequency signal analysis, in: Advances in Spectral Analysis and Array Processing, Prentice Hall, 1991, pp. 418–517, http://qspace.qu.edu.qa/handle/10576/10832.

[16] E. Bedrosian, A product theorem for Hilbert transforms, Proc. IEEE 51 (5) (1963) 868–869, http://dx.doi.org/10.1109/PROC.1963.2308.

[17] B. Boashash, Estimating and interpreting the instantaneous frequency of a signal, I: fundamentals, Proc. IEEE 80 (4) (1992) 520–538, http://dx.doi.org/10.1109/5.135376.

[18] T.-P. Jung, S. Makeig, C. Humphries, T.-W. Lee, M.J. Mckeown, V. Iragui, T.J. Sejnowski, Removing electroencephalographic artifacts by blind source separation, Psychophysiology 37 (2) (2000) 163–178, http://dx.doi.org/10.1111/1469-8986.3720163.

[19] M. Sun, S. Qian, X. Yan, S.B. Baumann, X.-G. Xia, R. Dahl, N. Ryan, R. Sclabassi, Localizing functional activity in the brain through time–frequency analysis and synthesis of the EEG, Proc. IEEE 84 (9) (1996) 1302–1311, http://dx.doi.org/10.1109/5.535248.

[20] A. Subasi, E. Ercelebi, A. Alkan, E. Koklukaya, Comparison of subspace-based methods with AR parametric methods in epileptic seizure detection, Comput. Biol. Med. 36 (2) (2006) 195–208, http://dx.doi.org/10.1016/j.compbiomed.2004.11.001.

[21] S. Altunay, Z. Telatar, O. Erogul, Epileptic EEG detection using the linear prediction error energy, Expert Syst. Appl. 37 (8) (2010) 5661–5665, http://dx.doi.org/10.1016/j.eswa.2010.02.045.

[22] V. Srinivasan, C. Eswaran, N. Sriraam, Approximate entropy-based epileptic EEG detection using artificial neural networks, IEEE Trans. Inf. Technol. Biomed. 11 (3) (2007) 288–295, http://dx.doi.org/10.1109/TITB.2006.884369.

[23] K. Polat, S. Gunes, Classification of epileptiform EEG using a hybrid system based on decision tree classifier and fast Fourier transform, Appl. Math. Comput. 187 (2) (2007) 1017–1026, http://dx.doi.org/10.1016/j.amc.2006.09.022.

[24] N. Kannathal, M.L. Choo, U.R. Acharya, P. Sadasivan, Entropies for detection of epilepsy in EEG, Comput. Methods Programs Biomed. 80 (3) (2005) 187–194, http://dx.doi.org/10.1016/j.cmpb.2005.06.012.

[25] B. Boashash, L. Boubchir, G. Azemi, A methodology for time–frequency image processing applied to the classification of non-stationary multichannel signals using instantaneous frequency descriptors with application to newborn EEG signals, EURASIP J. Adv. Signal Process. 2012 (1) (2012) 1–21, http://dx.doi.org/10.1186/1687-6180-2012-117.

[26] B. Boashash, G. Azemi, N.A. Khan, Principles of time–frequency feature extraction for change detection in non-stationary signals: applications to newborn EEG abnormality detection, Pattern Recognit. 48 (3) (2015) 616–627, http://dx.doi.org/10.1016/j.patcog.2014.08.016.

[27] S.S. Abeysekera, B. Boashash, Methods of signal classification using the images produced by the Wigner–Ville distribution, Pattern Recognit. Lett. 12 (11) (1991) 717–729, http://dx.doi.org/10.1016/0167-8655(91)90010-J.

[28] J. Terrien, C. Marque, G. Germain, Ridge extraction from the time–frequency representation (TFR) of signals based on an image processing approach: application to the analysis of uterine electromyogram AR TFR, IEEE Trans. Biomed. Eng. 55 (5) (2008) 1496–1503, http://dx.doi.org/10.1109/TBME.2008.918556.

[29] J.D. Martinez-Vargas, J.I. Godino-Llorente, G. Castellanos-Dominguez, Time–frequency based feature selection for discrimination of non-stationary biosignals, EURASIP J. Adv. Signal Process. 2012 (1) (2012) 1–18, http://dx.doi.org/10.1186/1687-6180-2012-219.

[30] P. Honeine, C. Richard, P. Flandrin, Time–frequency learning machines, IEEE Trans. Signal Process. 55 (7) (2007) 3930–3936, http://dx.doi.org/10.1109/TSP.2007.894252.

[31] S. Aviyente, E.M. Bernat, S.M. Malone, W.G. Iacono, Time–frequency data reduction for event related potentials: combining principal component analysis and matching pursuit, EURASIP J. Adv. Signal Process. 2010 (2010) 1–14, http://dx.doi.org/10.1155/2010/289571.

[32] B. Boashash, T. Ben-Jabeur, Design of a high-resolution separable-kernel quadratic TFD for improving newborn health outcomes using fetal movement detection, in: 11th International Conference on Information Science, Signal Processing and Their Applications, 2012, pp. 354–359.

[33] B. Boashash, Time-Frequency Signal Analysis and Processing: A Comprehensive Reference, Elsevier, Oxford, 2003.

[34] W.-k. Lu, Q. Zhang, Deconvolutive short-time Fourier transform spectrogram, IEEE Signal Process. Lett. 16 (7) (2009) 576–579, http://dx.doi.org/10.1109/LSP.2009.2020887.

[35] L.J. Stanković, A method for time–frequency analysis, IEEE Trans. Signal Process. 42 (1) (1994) 225–229, http://dx.doi.org/10.1109/78.258146.

[36] F. Hlawatsch, Interference terms in the Wigner distribution, Digit. Signal Process. 84 (1984) 363–367.

[37] B. Barkat, B. Boashash, A high-resolution quadratic time–frequency distribution for multicomponent signals analysis, IEEE Trans. Signal Process. 49 (10) (2001) 2232–2239, http://dx.doi.org/10.1109/78.950779.

[38] B. Boashash, V. Sucic, Resolution measure criteria for the objective assessment of the performance of quadratic time–frequency distributions, IEEE Trans. Signal Process. 51 (5) (2003) 1253–1263, http://dx.doi.org/10.1109/TSP.2003.810300.

[39] C. Yeh, A. Roebel, X. Rodet, Multiple fundamental frequency estimation and polyphony inference of polyphonic music signals, IEEE Trans. Audio Speech Lang. Process. 18 (6) (2010) 1116–1126, http://dx.doi.org/10.1109/TASL.2009.2030006.

[40] B. Boashash, Estimating and interpreting the instantaneous frequency of a signal, II: algorithms and applications, Proc. IEEE 80 (4) (1992) 540–568, http://dx.doi.org/10.1109/5.135378.
[41] V. Katkovnik, L.J. Stanković, Instantaneous frequency estimation using the Wigner distribution with varying and data-driven window length, IEEE Trans. Signal Process. 46 (9) (1998) 2315–2325, http://dx.doi.org/10.1109/78.709514.
[42] M.K. Emresoy, A. El-Jaroudi, Iterative instantaneous frequency estimation and adaptive matched spectrogram, Signal Process. 64 (2) (1998) 157–165, http://dx.doi.org/10.1016/S0165-1684(97)00183-7.
[43] P.L. Shui, H.Y. Shang, Y.B. Zhao, Instantaneous frequency estimation based on directionally smoothed pseudo-Wigner–Ville distribution bank, IET Radar Sonar Navig. 1 (4) (2007) 317–325, http://dx.doi.org/10.1049/rsn:20060123.
[44] L. Rankine, M. Mesbah, B. Boashash, IF estimation for multicomponent signals using image processing techniques in the time frequency domain, Signal Process. 87 (6) (2007) 1234–1250, http://dx.doi.org/10.1016/j.sigpro.2006.10.013.
[45] J. Lerga, V. Sucic, B. Boashash, An efficient algorithm for instantaneous frequency estimation of nonstationary multicomponent signals in low SNR, EURASIP J. Adv. Signal Process. 2011 (1) (2011) 1–16, http://dx.doi.org/10.1155/2011/725189.
[46] S. Dong, G. Azemi, B. Lingwood, P.B. Colditz, B. Boashash, Performance evaluation of multi-component instantaneous frequency estimation techniques for heart rate variability analysis, in: 11th International Conference on Information Science, Signal Processing and Their Applications, 2012, pp. 1211–1216.
[47] N.A. Khan, B. Boashash, Instantaneous frequency estimation of multicomponent nonstationary signals using multiview time–frequency distributions based on the adaptive fractional spectrogram, IEEE Signal Process. Lett. 20 (2) (2013) 157–160, http://dx.doi.org/10.1109/LSP.2012.2236088.
[48] I.M. Sobol, Simulation and the Monte Carlo Method, CRC Press, 2008.
[49] I. Shafi, J. Ahmad, S.I. Shah, F.M. Kashif, Computing deblurred time–frequency distributions using artificial neural networks, Circuits Syst. Signal Process. 27 (3) (2008) 277–294, http://dx.doi.org/10.1007/s00034-008-9027-x.
[50] L.J. Stanković, Analysis of noise in time–frequency distributions, IEEE Signal Process. Lett. 9 (9) (2002) 286–289, http://dx.doi.org/10.1109/LSP.2002.803409.
[51] K. Konstantinides, B. Natarajan, G.S. Yovanof, Noise estimation and filtering using block-based singular value decomposition, IEEE Trans. Image Process. 6 (3) (1997) 479–483, http://dx.doi.org/10.1109/83.557359.
[52] H. Hassanpour, A time–frequency approach for noise reduction, Digit. Signal Process. 18 (5) (2008) 728–738, http://dx.doi.org/10.1016/j.dsp.2007.09.014.
[53] R.G. Baraniuk, Wavelet soft-thresholding of time–frequency representations, in: IEEE International Conference on Image Processing, vol. 1, 1994, pp. 71–74, http://dx.doi.org/10.1109/ICIP.1994.413277.
[54] R.M. Haralick, S.R. Sternberg, X. Zhuang, Image analysis using mathematical morphology, IEEE Trans. Pattern Anal. Mach. Intell. 9 (4) (1987) 532–550, http://dx.doi.org/10.1109/TPAMI.1987.4767941.
[55] B. Dugnol, C. Fernández, G. Galiano, Wolf population counting by spectrogram image processing, Appl. Math. Comput. 186 (1) (2007) 820–830, http://dx.doi.org/10.1016/j.amc.2006.08.173.
[56] G.R. Arce, S.R. Hasan, Elimination of interference terms of the discrete Wigner distribution using nonlinear filtering, IEEE Trans. Signal Process. 48 (8) (2000) 2321–2331, http://dx.doi.org/10.1109/78.852013.
[57] L.J. Stanković, A measure of some time–frequency distributions concentration, Signal Process. 81 (3) (2001) 621–631, http://dx.doi.org/10.1016/S0165-1684(00)00236-X, special section on Digital Signal Processing for Multimedia.
[58] R.G. Baraniuk, D.L. Jones, A signal dependent time–frequency representation: fast algorithm for optimal kernel design, IEEE Trans. Signal Process. 42 (1) (1994) 134–146, http://dx.doi.org/10.1109/78.258128.
[59] D.L. Jones, R.G. Baraniuk, An adaptive optimal-kernel time–frequency representation, IEEE Trans. Signal Process. 43 (10) (1995) 2361–2371, http://dx.doi.org/10.1109/78.469854.
[60] T.-H. Sang, W. Williams, Renyi information and signal-dependent optimal kernel design, in: International Conference on Acoustics, Speech, and Signal Processing, vol. 2, 1995, pp. 997–1000, http://dx.doi.org/10.1109/ICASSP.1995.480344.
[61] D.L. Jones, T.W. Parks, A high resolution data-adaptive time–frequency representation, IEEE Trans. Acoust. Speech Signal Process. 38 (12) (1990) 2127–2135, http://dx.doi.org/10.1109/29.61539.
[62] M.J. Bastiaans, Comment on the t-class of time–frequency distributions: time-only kernels with amplitude estimation, J. Franklin Inst. 348 (9) (2011) 2670–2673, http://dx.doi.org/10.1016/j.jfranklin.2011.07.009.
[63] M.J. Bastiaans, T. Alieva, L.J. Stanković, On rotated time–frequency kernels, IEEE Signal Process. Lett. 9 (11) (2002) 378–381, http://dx.doi.org/10.1109/LSP.2002.805118.
[64] S. Choomchuay, K. Sihalath, An application of second derivative of gaussian filters in fingerprint image enhancement, in: 4th International Conference on Bioinformatics and Biomedical Engineering, 2010, pp. 1–4.
[65] Y. Liu, L. Mejias, Z. Li, Fast power line detection and localization using steerable filter for active UAV guidance, in: 12th International Society for Photogrammetry & Remote Sensing, Melbourne Convention and Exhibition Centre, Melbourne, VIC, 2012, http://eprints.qut.edu.au/53977/.
[66] M. Jacob, M. Unser, Design of steerable filters for feature detection using canny-like criteria, IEEE Trans. Pattern Anal. Mach. Intell. 26 (2004) 1007–1019, http://dx.doi.org/10.1109/TPAMI.2004.44.
[67] A.K. Ozdemir, O. Arikan, A high resolution time frequency representation with significantly reduced cross-terms, in: IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 2, 2000, pp. 693–696.
[68] D. Gabor, Theory of communication, Part 1: the analysis of information, Proc. Inst. Electr. Eng., Part 3, Radio Commun. Eng. 93 (26) (1946) 429–441, http://dx.doi.org/10.1049/ji-3-2.1946.0074.
[69] J.M. O'Toole, B. Boashash, Fast and memory-efficient algorithms for computing quadratic time–frequency distributions, Appl. Comput. Harmon. Anal. 35 (2) (2013) 350–358, http://dx.doi.org/10.1016/j.acha.2013.01.003.
[70] J. Lofhede, M. Thordstein, N. Löfgren, A. Flisberg, M. Rosa-Zurera, I. Kjellmer, K. Lindecrantz, Automatic classification of background EEG activity in healthy and sick neonates, J. Neural Eng. 7 (1) (2007) 1–12, http://dx.doi.org/10.1088/1741-2560/7/1/016007.
[71] V. Sucic, N. Saulig, B. Boashash, Estimating the number of components of a multicomponent nonstationary signal using the short-term time–frequency Renyi entropy, EURASIP J. Adv. Signal Process. 2011 (1) (2011) 1–11, http://dx.doi.org/10.1186/1687-6180-2011-125.
[72] R.G. Baraniuk, P. Flandrin, A.J.E.M. Janssen, O.J.J. Michel, Measuring time–frequency information content using the Renyi entropies, IEEE Trans. Inf. Theory 47 (4) (2001) 1391–1409, http://dx.doi.org/10.1109/18.923723.
[73] L. Rankine, M. Mesbah, B. Boashash, A matching pursuit-based signal complexity measure for the analysis of newborn EEG, Med. Biol. Eng. Comput. 45 (3) (2007) 251–260, http://dx.doi.org/10.1007/s11517-006-0143-0.
[74] T. Fawcett, An introduction to ROC analysis, Pattern Recognit. Lett. 27 (8) (2006) 861–874, http://dx.doi.org/10.1016/j.patrec.2005.10.010.
[75] N.J. Stevenson, M. Mesbah, G.B. Boylan, P.B. Colditz, B. Boashash, A nonlinear model of newborn EEG with nonstationary inputs, Ann. Biomed. Eng. 38 (9) (2010) 3010–3021, http://dx.doi.org/10.1007/s10439-010-0041-3.
[76] A.M. Sayeed, D.L. Jones, Optimal kernels for nonstationary spectral estimation, IEEE Trans. Signal Process. 43 (2) (1995) 478–491, http://dx.doi.org/10.1109/78.348130.
[77] F. Auger, P. Flandrin, Y.-T. Lin, S. McLaughlin, S. Meignen, T. Oberlin, H.-T. Wu, Time–frequency reassignment and synchrosqueezing: an overview, IEEE Signal Process. Mag. 30 (6) (2013) 32–41, http://dx.doi.org/10.1109/MSP.2013.2265316.
[78] A. Loza, N. Canagarajah, D. Bull, Region feature-based segmentation of time–frequency images, in: International Workshop on Systems, Signals and Image Processing, Ambient Multimedia, Poznan, 2004, pp. 375–378.
[79] N. Whitelonis, H. Ling, Radar signature analysis using a joint time–frequency distribution based on compressed sensing, IEEE Trans. Antennas Propag. 62 (2) (2014) 755–763, http://dx.doi.org/10.1109/TAP.2013.2291893.
[80] T.H. Cormen, C. Stein, R.L. Rivest, C.E. Leiserson, Introduction to Algorithms, 2nd edition, McGraw-Hill Higher Education, 2001.
[81] M. Hansson-Sandsten, Multitaper Wigner and Choi–Williams distributions with predetermined Doppler–lag bandwidth and sidelobe suppression, in: Fourier Related Transforms for Non-Stationary Signals, Signal Process. 91 (6) (2011) 1457–1465, http://dx.doi.org/10.1016/j.sigpro.2010.10.010.
[82] B. Boashash, Algorithms for time–frequency signal analysis, in: Time–Frequency Signal Analysis – Methods & Applications, Longman-Cheshire, John Wiley Halsted Press, Melbourne, 1992, pp. 163–181, http://qspace.qu.edu.qa/bitstream/handle/10576/10811/Boashash-Reilly_1992_Longman_Cheshire_TFSA-book_Algorithms-4-TFSA.pdf?sequence=1.

Boualem Boashash (IEEE Fellow '99') is a Scholar, Professor and Senior Academic with experience in 5 leading Universities in France and Australia. He has published over 500 technical publications, including over 100 journal publications covering Engineering, Applied Mathematics and Bio-medicine. He pioneered the field of Time-Frequency Signal Processing for which he published the most comprehensive book and most powerful software package. Among many initiatives, he founded ISSPA, a leading conference since 1985. After a previous appointment in the UAE as Dean of Engineering, he moved to Qatar University as a Research Professor. His work was cited over 10,000 times.

Nabeel Ali Khan received his Ph.D. from Mohammad Ali Jinnah University, Islamabad, Pakistan, in 2010. From June 2008 to August 2012, he was a faculty member in Federal Urdu University, Islamabad and senior design engineer at Center for Advanced Research in Engineering. In 2012, he joined the Department of Electrical Engineering, Qatar University, Doha, Qatar as a Post-Doctoral Fellow, working on automatic classification of EEG abnormalities and EEG artifact removal. His research interests include time–frequency signal analysis, software defined radios, and pattern recognition.

Taoufik Ben-Jabeur received his M.S. and Ph.D. degrees in Computer Science and Signal Processing from University of Paris-Descartes, France, in 2005 and 2009, respectively. In 2010, he joined the University of Reims, France, as a Postdoc in the area of wireless multicarrier communications and sensor networks. From May 2011 to May 2014, he was with Qatar University as a Post-Doctoral Fellow. Starting from June 2014, he is a Research Scientist at Qatar Mobility Innovations Center (QMIC). His research interests include wireless sensor networks system, multicarrier communications, RFID system and time frequency analysis.